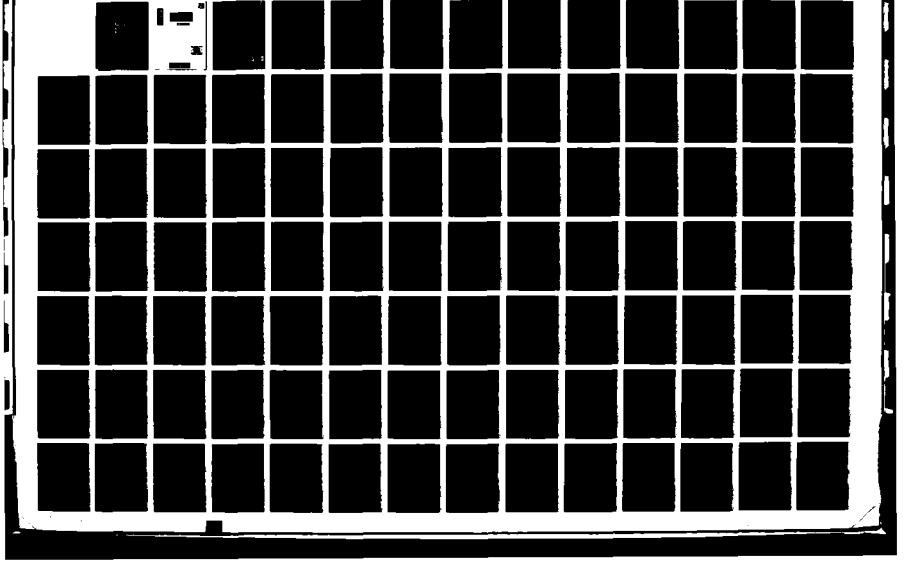
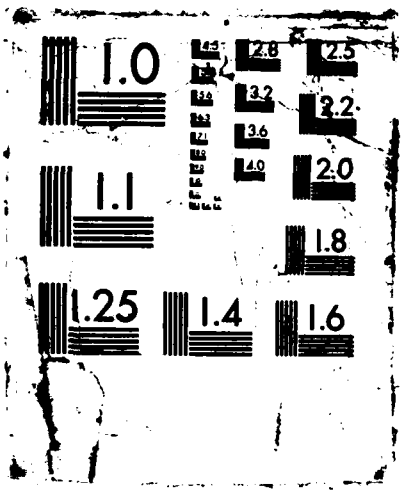
AD-A193 362 CAPITALIZING ON EXPERIENCE WITH INTELLIGENCE GATEWAY SOFTWARE(U) LOGISTICS MANAGEMENT INST BETHESDA MD C W SHOCKLEY JAN 88 LMI-DL604R1 DTIC-TR-88/7 1/2

UNCLASSIFIED MDA903-85-C-0139 F/G 12/5 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

CAPITALIZING ON EXPERIENCE WITH INTELLIGENT GATEWAY SOFTWARE

Report DL604R1

January 1988

Cynthia W. Shockley

Prepared pursuant to Department of Defense Contract MDA903-85-C-0139 under Task DL604 sponsored by the Defense Technical Information Center. The views expressed here are those of the Logistics Management Institute at the time of issue but not necessarily those of the Department of Defense. Permission to quote or reproduce any part must – except for Government purposes – be obtained from the Logistics Management Institute.

LOGISTICS MANAGEMENT INSTITUTE
6400 Goldsboro Road
Bethesda, Maryland 20817-5886

Executive Summary

CAPITALIZING ON EXPERIENCE WITH INTELLIGENT GATEWAY SOFTWARE

Intelligent gateway software helps database users search, retrieve, and analyze information from many different data systems, even when severe hardware and software incompatibilities exist. The Defense Technical Information Center (DTIC) and the National Library of Medicine (NLM) each have developed gateway software. DTIC has two gateway packages – the Defense Gateway Information System (DGIS) and Search Menu-Aided Easy Searching Through Relevant Options (SearchMAESTRO). NLM also has two – Micro-Chemical Substances Information Network (Micro-CSIN) and Grateful MED.

Each of these gateways has its strengths and weaknesses. DTIC and NLM experience with the gateways, along with gateway software developments in the commercial sector and in academia, reveal opportunities for improvements.

We find that:

- DTIC's DGIS gateway software excels in processing the results of a database search, but offers the user little help in devising search strategies for the databases selected. DTIC's SearchMAESTRO provides easier database searching, particularly for novice users. DTIC software gives no guidance for choosing between the two gateways according to individual need and experience.
- With NLM's Micro-CSIN, both experienced and novice users can search databases, using "scripts" which guide the users in defining their search strategies. Although the user interface is unwieldy, Micro-CSIN provides the user with excellent assistance in composing a search strategy and selecting relevant databases. Despite the limitation of being able to search only two databases, Grateful MED is particularly valuable in enabling a user to review and incorporate terms and names from lists of controlled vocabularies into a search strategy. Initial plans for NLM's Advanced, Biomedical Information Databand Exchange (ABIDE) gateway structure, whose purpose is to improve further the dissemination of information in biotechnology, are technically sound. As yet, NLM has not adapted present gateway offerings into the ABIDE structure.

- Several gateway developments in the commercial and academic sectors, such as the object-oriented gateway environment which supports retrieval from databases of dissimilar contents and expert systems which assist users in constructing search strategies, hold promise for both DTIC and NLM gateway enhancements.

We recommend that:

- DTIC enhance DGIS with the capabilities similar to those in NLM's Micro-CSIN and Grateful MED that enable both experienced and novice users to structure search strategies and select suitable databases. DTIC should adapt DGIS so that search terms used for one database can be used in several databases. Such DGIS improvements should enable DTIC to use the one gateway to meet the information needs of all users.

- NLM improve the "user-friendliness" of Micro-CSIN by incorporating online and "window" prompting techniques. Micro-CSIN should also be modified to incorporate controlled vocabulary terms and names into a search strategy. For the long term, NLM should decide how Micro-CSIN and Grateful MED will be incorporated into the ABIDE gateway framework.

- Both DTIC and NLM adopt the structure of an object-oriented gateway environment. They should also develop expert systems to aid users in creating more effective search strategies and consider cooperative funding arrangements for future gateway development.

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

CONTENTS (Continued)

CONTENTS (Continued)

CONTENTS (Continued)

TABLES

FIGURES

FIGURES (Continued)

SECTION 1

INTRODUCTION

In other words, then, if a machine is expected to be infallible, it cannot also be intelligent.

-Alan Turing[1]

Intelligent gateway software helps database users search, retrieve, and analyze information from many different data systems, despite hardware and software incompatibilities among the systems. The Defense Technical Information Center (DTIC) and the National Library of Medicine (NLM) have each developed gateway software. DTIC and NLM now wish to see how their respective development activities can be improved, how gateway software from the commercial sector and from academia can be used to enhance their gateways, and how each may benefit from the accomplishments of the other.

We also considered adapting gateway software developed by the Lawrence Livermore National Laboratory (LLNL) as a possible front-end to NLM's Medical Literature Analysis and Retrieval System (MEDLARS). The LLNL software, known as the Technology Information System (TIS) Intelligent Gateway software, is now in use at DTIC and is the basis for that organization's Defense Gateway Information System (DGIS).

Another DTIC gateway, Search Menu-Aided Easy Searching Through Relevant Options (SearchMAESTRO), developed by Telebase Systems for DTIC, is also a part of DTIC's overall gateway capabilities. Optional additions to the DTIC gateway include the features available in the Micro-Chemical Substances Information Network (Micro-CSIN) and Grateful MED software developed at NLM.

The four DTIC and NLM gateway packages are compared based on the following categories:

- Gateway features -- presearch strategy development and database selection, search execution, and post-processing of search results

[1]Hodges, Andrew. *Alan Turing: The Enigma.* New York: Simon and Schuster, Inc., 1983.

- Gateway utilities, e.g., file-handling capabilities
- Gateway support for users
- Gateway functionality, e.g., operating system and use of menus versus command structures.

We also consider the functional highlights and technical strengths of several other gateway packages offered as research tools in academia or as products in the marketplace. With this comparative foundation in place, the last chapter contains our recommendations to both DTIC and NLM for further gateway developments.

1.1 GATEWAY: A DEFINITION

A data gateway is typically a software package that facilitates searching, retrieval, and analysis of information or data from systems that are dissimilar in hardware and software. In essence, data gateways all have the same goal: to make the complexities of online information searching and retrieval transparent to users. An intelligent gateway, which is what this report focuses on, provides more than a simple automatic log-on to a database. An intelligent gateway handles a transparent log-on by routing an information retrieval question automatically to an appropriate system without requiring the user to know telephone numbers, protocols, and passwords. The gateway translates user-prompted queries into a form that can be read by one or more database retrieval systems. It directs a user to one or more multiple online systems and ultimately to one or more databases within the systems. In more and more cases, a gateway can support downloading and continued analysis of the retrieved information.

1.2 THE NEED FOR GATEWAYS

In recent years, the number of publicly available online information systems and databases has increased dramatically. It is estimated that more than 2,500 databases are available on 300 or more systems.[2] The systems are used both by end-users who need the information and by search intermediaries who typically perform information searches for others. When fewer than 100 databases and fewer than 5 systems were available, it was possible for one person to stay reasonably current and maintain searching skills on all of them. Those days are gone.

---

[2]*Computer-Readable Databases: A Directory and Data Sourcebook*. Williams, M. E., L. Lannom, and C. Robins, ed. Chicago, Ill.: American Library Association, 1984.

Although trends in both hardware and software technology are providing users of information analysis and support systems with significantly greater capabilities, the proliferation of systems that contain the data – whether textual, numeric, pictorial, graphical, or some combination – has created an environment in which the data are separately owned, maintained, and stored on a variety of media. Users of the data – whether they are trained search intermediaries or end-users – are faced with the formidable task of mastering several database-access and retrieval languages, combining the retrieved data into unified sets, and integrating the results further for subsequent review and reporting. The result is a classic Tower of Babel syndrome, where construction of a viable and coherent information query across multiple database systems by an individual has become almost impossible. Development of gateway software has evolved as one approach toward solving the problem.

1.3 MAJOR GATEWAY COMPONENTS

Research in gateway development began in the mid-1970s, and operational systems became available in the early 1980s. Most gateway software has reached the point of addressing, in some capacity, one or more of the components described below. Typically, an organization that chose to develop a gateway began by concentrating on one or two of these components. Consequently, any gateway in the marketplace has had a special strength in one or more of these components. Now many gateway packages are being augmented to address all of the component areas more fully.

1.3.1 Gateway Features

1.3.1.1 Conducting Presearch Activities and Selecting Databases

Presearch and selection of databases constitute what is probably the newest area of research in gateway development. To successfully address this feature, a gateway must help the user to formulate a search strategy that works well with one or more selected databases. Gateways have also begun directing users to the databases that may be most useful to them, in the context of the search strategy.

1.3.1.2 Executing the Search Strategy

Carrying out the search strategy has meant, primarily, automation of the dial-up, log-on, and submission of a translated search query steps to a database. In some

instances, a gateway can continue to translate and submit the original search strategy to other databases as each search is completed.

1.3.1.3 Performing Post-Processing on the Results of a Search Session

Information retrieved from databases often requires continued analysis – called post-processing – to become more valuable to the user. Post-processing may, for example, eliminate duplicate citations in downloaded results or suggest additional search terms that could further enhance the search strategy.

1.3.2 Gateway Utilities

Gateway utilities support such subsidiary software or features as accounting, electronic mail (EM), word processing, and file editing.

1.3.3 Gateway Support

This component concerns interaction with and support for the user, e.g., help subsystems.

1.3.4 Gateway Functionality

Functionality involves a review of the versatility of the operating system and transportability of software. Use of menu structures, command-mode structures, or both, as well as access capabilities, are also considered within this component.

1.4 GATEWAY ENVIRONMENTS

Gateway software can be provided in one of two main ways: a multi-user mainframe or minicomputer, or a single-user microcomputer. The central-server, multi-user approach is exemplified by TIS from LLNL, its derivative DGIS from DTIC, and EasyNet (offered by Telebase, Inc.), which is available as Search-MAESTRO from DTIC.

Microcomputer gateway software is divided further into two categories: packages that interface to a single-host system and those that interface to multiple-host systems. In either category, these microcomputer packages can be dedicated to searching functions only or can take on other functions by the addition of, for example, text editors (or word processing), or accounting software. The single-host gateway microcomputer software is often marketed and distributed by the host

database vendor to encourage use of its databases. Some examples of this type of software are: DIALOGLINK (or OneSearch), Search Helper, MicroDISCLOSURE, Searchware, Grateful MED, DunsPlus, and Wilsearch. Examples of microcomputer software packages that enable the user to interface with multiple host systems are such products as Pro-Search, Micro-CSIN, Sci-Mate, and PC/Net-Link.

Two of the gateways evaluated are large, multi-user, mainframe-environment gateways (DGIS and SearchMAESTRO available from DTIC), and two are single-user microcomputer-environment packages (Micro-CSIN and Grateful MED), offered by NLM. Thus far, these two organizations have approached gateway development from a centralized-versus-decentralized approach. However, NLM had originally begun CSIN development on a minicomputer, and DTIC is considering a PC-based version of DGIS. In this report, DGIS is compared with Micro-CSIN, and SearchMAESTRO with Grateful MED. This approach was chosen because the gateways in the two groups are functionally similar; that is, DGIS and Micro-CSIN are geared to support a more sophisticated user group, typically intermediaries, and SearchMAESTRO and Grateful MED are designed to retrieve database search results quickly, with limited interaction, usually by a novice or inexperienced end-user. As a consequence of comparing gateway features, utilities, support, and function developed for different computer environments, one package may not address a given area in a comparable manner to other gateway environments. In this report, however, approaches to gateway software development are considered independently of their hardware environment when recommendations for continued gateway development are presented to DTIC and NLM.

SECTION 2

DoD GATEWAY INFORMATION SYSTEM

DoD's Research, Development, Test, and Evaluation (RDT&E) community requires rapid and easy access to scientific and technical information (STI) pertinent to its mission. The STI is contained in a number of databases maintained in the Federal and commercial sectors. DTIC has as its mission the acquisition, storage, retrieval, and dissemination of STI to support the RDT&E community mission. To provide this community with a modern tool for identifying, accessing, and retrieving information from these databases, DTIC is developing a gateway system.

This is the DGIS, a low-level artificial intelligence (AI)-like system that operates on BSD UNIX and is called the Intelligent Gateway Processor (IGP). IGP, developed by the Department of Energy's (DoE's) LLNL TIS, enables the user to extract information from local and remote databases, and to reformat and transfer that information. The main menu for the DGIS is displayed in Figure 2-1.

Initial DGIS development, testing, and evaluation was funded by DTIC and performed at LLNL. This prototype DGIS is now undergoing test and evaluation within the DoD community. DTIC plans to stabilize a version of the DGIS and offer it as a standard service in October 1988. Development of a classified gateway is also underway.

Another gateway software package offered is SearchMAESTRO. Search-MAESTRO is described separately in Section 3. Although use of SearchMAESTRO can be preceded by use of DGIS, they are separate entities.

2.1 GATEWAY FEATURES

2.1.1 Conducting Presearch Activities and Selecting Databases

At present, a user who accesses the DGIS must request the SearchMAESTRO option in order to be offered assistance with presearch strategy development and database selection. DGIS does not now provide a search strategy set-up or search

command translator to help a user execute an information search request. A user who is unprepared to construct search strategies in the DGIS "native mode" (or mode that places the user at the entry of the database without the guidance provided by a gateway) has the option of using SearchMAESTRO as a gateway to many of the same databases that are accessible via DGIS. Two enhancement projects are underway at DTIC – development of an online resource directory and development of the common command language (CCL) – which will broaden DGIS capabilities in these feature areas (see Sections 7.1.1 and 7.1.2).

```
                    Welcome to the DoD Gateway Information System

>>>>>>>>>>>>INFORMATION TRANSFER MODULES
1       directory                   DGIS Directory of Online Resources.
2       communicate                 Connect to Information resources and people.
3       process                     Information product tailoring.

 >>>>>>>>>>>>>INFORMATION UTILITIES
4       em                          Electronic Mail.
5       files                       File operations.

>>>>>>>>>>>>>SUPPORT INFORMATION
6       help                        Description of features.
7       users                       DGIS registered users.
8       info                        DGIS news and information.
9       dticlog                     DGIS full text retrieval.

DGIS HOTLINE NUMBERS:               (800) 255-GATE (Outside Virginia)
                                    (703) 276-8182 (Inside Virginia)

or send questions via DGIS EM to 'dgishelp'.

Enter a menu number, a command, "b" to backup, "t" for top, or "e" to end:
*
```

FIG. 2-1. DGIS MAIN MENU

2.1.2 Executing the Search Strategy

DGIS and SearchMAESTRO support search strategies in separate ways. In DGIS, the user is actively involved in the interaction with the database as searching takes place. The user sees the results of the search session and can continue to refine and modify the search strategy to improve the results. Use of SearchMAESTRO, on the other hand, has no interactive exchange between the user and the system.

Database searching with SearchMAESTRO takes place under the direction of the gateway without intervention by the user.

DGIS can provide access to most computer systems available on Tymnet, Telenet, or by direct dialing. In cases of frequently accessed vendor systems, such as DIALOG and the National Aeronautics and Space Administration's Remote Console (NASA/RECON) system, automatic connections can be selected and the user is logged-on directly. The user selects the "communicate" option from the DGIS main menu to begin a search on a selected database. Figure 2-2 shows the screen as the user sees it.

```
2  COMMUNICATIONS

DGIS will automatically connect you to a wide range of remote
information systems and to other people online the DGIS.  For
information systems, you must have already registered with
these systems and have provided DGIS your access passwords.

>>>>>>>>>>>>> ASSISTED SEARCHING
2.1  assist      Search interfaces menu.

>>>>>>>>>>>>> NON-ASSISTED SEARCHING
2.2  connect     Auto-login to remote systems.
2.3  systems     List of systems with auto access.
2.4  dial        Unassisted dial into other systems.

>>>>>>>>>>>>> OTHER COMMUNICATIONS
2.5  people      Communicate interactively with DGIS users.

Enter a menu number, a command, "b" to back up, "t" for top, or "e" to end:
*
```

FIG. 2-2. DGIS COMMUNICATIONS MENU

SearchMAESTRO, the first interface available under the "ASSISTED SEARCHING" option, is discussed in Section 3.

A user can review systems and databases that are available and accessible by entering a "2.2: connect – Auto-login to remote systems" command at the prompt (*). If the user chooses access to multi-type information systems, under the "2.3: systems – List of systems with auto access" option (which are predominantly bibliographic, with some factual/numeric, directory, and full-text databases included), the screen that is presented is as shown in Figure 2-3.

2.3 systems

>>>>LIST OF SYSTEMS TO WHICH DGIS CAN CONNECT AUTOMATICALLY>>>>

These are systems for which DGIS has connection agents already. A user requires a password and permission from DGIS in order to actually connect to them. To find out what systems you have access to, use the 'connect' command.

maestro	SearchMAESTRO - menu driven search interface.

MULTI-TYPE INFORMATION SYSTEMS

brs	BRS/Search Service.
cas-online	STN/Chemical Abstracts System.
compuserve	
dial	Dial to a system with a user-entered phone number.
tymnet	Connect to the TYMNET value-added network.
telenet	Connect to the TELENET value-added network.

Enter a <RETURN> to continue.
*

Notes: BRS = Bibliographic Retrieval Service; STN = Scientific and Technical (Information) Network.

FIG. 2-3. DGIS SYSTEMS MENU SELECTED AS COMMUNICATIONS OPTION

A user can select from any of these options – if accounts have been established with the vendor – and then be connected to the database or service by DGIS. Once logged onto the system, the user must be able to construct and execute a search strategy using the specific commands germane to that database vendor or database. A search strategy created using a file editor may also be used.

DGIS can also connect a user to any computer that can be dialed through commercial telephone lines. This option is available through selection of "2.4: dial – Unassisted dial into other systems" from the main communications menu (see Figure 2-2). The user can enter the telephone number for the desired system, along with such other pertinent information as baud rate, parity, and duplex. The DGIS can then establish the communications link with the system, and the user can log onto the computer.

It is interesting to note that another iteration of the SearchMAESTRO gateway – IQUEST – is available to users who have established a CompuServe account. All of the capabilities offered by the basic EasyNet service that address the

needs of a more general business and professional user, but not necessarily DoD population, are available.

2.1.3 Performing Post-Processing on the Results of a Search Session

DoE recognized the need for bibliographic information post-processing and developed several options available to the DoE/RECON user via TIS. DTIC also funded development of post-processing features on TIS. This library of post-processing routines for numeric and bibliographic data was available on TIS software and was imported to the DGIS project. This function, as adapted by DTIC, is called "Information product tailoring" on the DGIS main menu (see Figure 2-1). The menu for bibliographic post-processing, if selected by the user, appears as shown in Figure 2-4.

```
3.1  BIBLIOGRAPHIC PROCESSING

Bibliographic processing will allow you to standardize your
downloaded files and merge them into a single working file.
You will then be able to process that file to result in a
finished bibliography tailored to your needs.

3.1.1   translate    Translate citations to a standard format.
3.1.2   merge        Merge translated files into a single file.
3.1.3   duplicate    Eliminate duplicate citations.
3.1.4   analysis     GO TO CITATION ANALYSIS MENU.
3.1.5   prepare      GO TO BIBLIOGRAPHIC PREPARATION MENU.
3.1.6   news         What's new on Bibliographic Processing.
3.1.7   display      Display the contents of a file.
3.1.8   automatic    Translate, duplicate elimination, sort
                     (year, author, title) and print final.

Enter a menu number, a command, "b" to backup, "t" for top, or "e" to end:
*
```

FIG. 2-4. DGIS BIBLIOGRAPHIC RECORDS POST-PROCESSING MENU

The DGIS bibliographic post-processing features enable the user to analyze, review, and rearrange downloaded bibliographic citations. At present, numeric data

are not handled in these post-processing routines. These are the post-processing features for bibliographic records:

- Translate downloaded files into a standard bibliographic format

  For DGIS to process downloaded bibliographic citations, they must all be in a standard format, i.e., all fields must be tagged the same. The DGIS citation translator converts the citations into this standard format irrespective of the database of origin. If the DGIS translator does not recognize a field, the original form is retained and can be used.

- Merge several files into a single file

  If several files from various database vendors have been downloaded and translated, they should be merged to enable the user to review unique citations. Up to five files can be merged at a time. As a result, a new, merged file is created, and the various files it comprises are left intact in the user's DGIS file directory.

- Eliminate duplicate citations

  After merging records from several different databases, the user may find duplicates in the merged file. With the DGIS duplicate-elimination feature, the user can select the fields that are to be compared. The DGIS default selection is to compare the year, author, and title fields of the citations. The present feature provides a character string matching capability only.

- Cross-correlate fields

  The DGIS cross-correlate – or citation analysis – feature provides: statistical analysis of the frequency of fields, frequency counts in two fields, cross-correlation of frequency counts in two fields, and plotting of citations for a given year in a file.

- Prepare an index or bibliography

  This feature enables the user to examine each citation in sequential order and decide whether it is pertinent to the user's information need. As each citation is displayed, the user decides whether to retain or discard it. With another feature, the user can perform a multi-level sort of citations – e.g., by author and within the author field in reverse chronological order of the year of publication. The user can also prepare an index that sorts the entries in a file by a specified field. Finally, the user can prepare a bibliography by reformatting the citations. This procedure deletes such extraneous information as field labels from the citations.

- Automatic post-processing feature

  This automatic feature enables a user to perform several post-processing options in a single step. The DGIS automatically translates a file into the standard format; eliminates any duplicates; sorts the citations by year, author, and title; and arranges the citations in a new format. This feature is now limited to files whose citations must all come from the same system. Files from different systems can also be chosen for automatic processing after they have been merged. Unlike the use of individual commands – such as "duplicate" or "translate" as described above – if the automatic mode is invoked, the intermediate files created during the processing are deleted by DGIS.

These post-processing routines can be effective in assisting a user – whether intermediary or end-user – to better manage and use the search results obtained from several databases.

2.2 GATEWAY UTILITIES

Gateway utilities, as defined in Section 1.3.2, provided by the DGIS gateway include EM, file handling and editing capabilities, and establishment of subdirectories.

2.2.1 Electronic Mail

EM enables a DGIS user to send and receive mail from users of the same computer system and of other computer systems via a user's terminal that can gain access to DGIS. One major feature of EM is the ability to search and display information about DGIS users and user groups. Private mail groups can also be created by the user if regular messages are sent to a group of DGIS users.

2.2.2 File Handling and Editing Capabilities

Almost every DGIS operation results in the creation or modification of a file. Data captured from remote database systems are stored in DGIS files. Post-processing activities in these files create more files. EM messages can be stored in a file. A number of tools are available to store and manipulate these files. Files can be copied, moved, renamed, deleted, and edited. Two editing choices are available – Visual Editor (Vi) and a line editor (Ex). Vi is the UNIX visual or full-screen editor. Vi offers word processing features; it can insert, delete, or move text. Vi can modify downloaded information, edit mail messages, or create new files. Ex is

the line editor. The user may work on only one line of text (or group of lines) at a time. The cursor cannot be moved around the screen.

DGIS enables a user to restrict or open access to the files and directories to other DGIS users. DGIS automatically restricts newly created files and directories so that no other user can display or copy them. However, a user can alter these initial settings.

Users can also create subdirectories for file handling. Placement of files in subdirectories enables a user to group similar files to reduce the large number of separate files that can accumulate in the "home" or log-on directory. When a user establishes a subdirectory, a tree structure (or hierarchical directory) is used. Each subdirectory can be grouped by content, source, etc. In addition to creating subdirectories, a user may delete them; the user can also move items among subdirectories.

2.2.3 Accounting

DGIS and SearchMAESTRO establish master accounts for the database services that are offered. Both gateways can then provide users with individual invoices and account statements. In the future, DGIS activities will be charged to a user's National Technical Information Service (NTIS) account.

2.3 GATEWAY SUPPORT

Features specifically available on DGIS are discussed below. A Gateway User Support and Training Office (GUSTO) has also been established. GUSTO staff members provide a hotline service that users can call when they have problems. The staff members identify the source of the problem (for instance, the gateway, the user's terminal, a telecommunications link, or the remote system) and take action to have the problem resolved. Users may also contact the GUSTO staff via EM, which is available on DGIS.

2.3.1 Online Help

Limited context-sensitive online help is now available to the DGIS user. If a user types the command "help" at a prompt, a screen of available DGIS commands with a brief explanation is provided. For example:

"copy	Copies a file under a new name in your account
count	Counts number of times term appears in a field
cross	Correlate frequency counts for two fields
delete	Deletes a file in your account."

The same command listing is made available if a user selects "6: help – Description of features" from the main DGIS menu.

2.3.2 Registered Users Directory

This gateway option provides an online list of DGIS users in one of two ways. First, all users are listed alphabetically by their user identification. Second, a search can be performed on the registered users' file. A string of characters, which can be case-insensitive, can be entered for comparison in any of the fields, e.g., last name or organization. Retrieved user records are then displayed.

2.3.3 DGIS News and Information

Recent events and system updates are provided periodically under this gateway function. More typically, these updates appear as system messages when a user logs onto the DGIS.

2.4 GATEWAY FUNCTIONALITY

The IGP toolbox provides a two-level interface for the user to DGIS. DGIS, thus, provides the user with two levels of information handling – one casual (menus) and one knowledgeable (commands).

2.4.1 Menus

The menu mode offers all the DGIS options in a logical progression of menus. Menu access was considered an important issue during the DGIS design phase. Going up and down through menus to get from one capability to another was judged tedious. The concept of accessing any menu module from any area in the menu structure was incorporated. DTIC also determined that menu-use continuity was

important, and a way to escape a given gateway process must always be evident to the user.

2.4.2 Command Mode

Every DGIS menu option can also be executed as a command and can be used at any point within DGIS. This capability enables an experienced user to move more quickly around the DGIS features without being forced into the slower pace imposed by a "menu only" selection.

An enhanced DGIS command feature is the Command Pattern Search System (COPS), which was conceived, designed, and developed by a DGIS technical staff member to enable users to use truncated DGIS commands to speed activity and decrease key-stroking. A command may be truncated at any length, down to a command entry that remains unique. For example, the system command "communicate" may be entered as: "communicat," "communic," "commun," or "com" – but not "co," which could be confused with the system commands "connect" and "count." In addition, COPS tolerates faulty entries and adjusts them. For example, if a user accidentally enters "delate," COPS responds with: "Do you mean 'delete'? (y/n)." If the user answers "yes," the command is invoked. If "no," the user is retained at the last point of session activity with the reappearance of the menu prompt.

2.4.3 Operating System and Database Management System

The DGIS prototype is running on a VAX 11/780, using the UNIX operating system, the INGRES database management software, and the PROgramming LOGic (PROLOG) interpreter package. The DGIS software has been ported to a Pyramid 98X, an Elxsi 6800, a Gould 6050, and Sun Workstations for benchmarking and performance evaluation. Based on the results of the performance evaluation, a hardware configuration for a production system will be acquired. The production configuration may consist of several machines networked together.

2.4.4 Multiple Operations

DGIS enables the user to carry on several database retrieval operations at one time. This is done by placing one or more retrieval and downloading sessions in a background mode. This capability, unique to DGIS among the four gateways reviewed, is more complex than other DGIS operations but is particularly useful

when several remote database systems are being accessed. This process requires that the user leave the DGIS menus and work directly with the UNIX operating system.

2.4.5 Support of Various Types of Terminals

DGIS supports a large number of terminal types for access to its prototype system. Virtually any terminal or personal computer (PC) with telecommunications capability can access the DGIS.

2.4.6 Telecommunications Access

To access DGIS, a user can use a modem that is compatible with the selected terminal and is capable of 300, 1200, or 2400-baud communications. Modems must support Bell 212A communications protocol unless they communicate at 300 or 2400-baud. Although not shown as a menu selection, the Kermit file transfer and communications program is available.

Communications via DGIS to outside databases, Tymnet, and the Defense Data Network are supported by high-speed modems associated with the computer that supports the DGIS.

2.5 GATEWAY SUMMARY

DGIS has become a well-known prototype gateway, which has benefited from the work done by the LLNL and Control Data Corporation on behalf of DoE and DTIC. DGIS is on the verge of moving from a prototype environment to an operational one. It provides access to a wide range of Federal and commercial databases. DGIS enables the user to interact directly with the databases in building and executing a search strategy. Such direct interaction is typically conducted by an experienced searcher – not a novice user – for whom the capabilities of an alternative gateway, such as SearchMAESTRO, are more appropriate. This DGIS gateway excels at performing post-processing and continued editing of bibliographic retrieval results. In addition, as a large multi-user based system, it supports a useful EM feature.

A number of projects, now in various stages of evolution, are expected to enhance DGIS capabilities (see Section 7.1). The Directory of Resources, improved post-processing, Common Command Language System (CCLS), and others will make DGIS more effective as a gateway.

SECTION 3

SEARCHMAESTRO

Another DTIC gateway software package offered is SearchMAESTRO. This is a gateway interface based on the EasyNet system, tailored for a DoD-oriented service by Telebase Systems, Inc. Telebase developed the EasyNet service in 1984 under sponsorship of the National Federation of Abstracting and Information Services (NFAIS).

DoD end-users have the option of accessing SearchMAESTRO by first logging on to DGIS or by dialing directly to the Telebase computer in Pennsylvania. The idea behind SearchMAESTRO is to make available a common approach, easy-to-use, external interface for DoD end-users to search diverse topic-relevant databases – including the Defense RDT&E Online System (DROLS) Technical Reports (TR) database – for comprehensive information, in a coordinated, uniform manner. In addition to DROLS, the DoD end-user, through SearchMAESTRO, has a common-approach access to over 800 databases. The user accesses SearchMAESTRO from the DGIS main menu by requesting the "communicate" feature (see Figure 2-1). When the user has selected this DGIS option and logged into SearchMAESTRO, the main menu is as shown in Figure 3-1.

PRESS	TO SELECT
1	See information about SearchMAESTRO
2	See SearchMAESTRO pricing
3	Leave SearchMAESTRO
H	Help

-->

FIG. 3-1. SEARCHMAESTRO MAIN MENU

3.1 GATEWAY FEATURES

3.1.1 Conducting Presearch Activities and Selecting Databases

At present, a user who accesses DGIS must request the SearchMAESTRO option to be offered assistance in presearch strategy development and database selection. DGIS does not provide a search strategy set-up or search command translator to assist a user in executing an information search request. If a user is unprepared to construct search strategies in the DGIS "native mode," the option of using SearchMAESTRO as the gateway to many of the same databases that are accessible via DGIS is available. Two enhancement projects are underway at DTIC – development of an online resource directory and development of the CCL – which will broaden DGIS capabilities in these feature areas (see Section 7.1.1 and Section 7.1.2).

SearchMAESTRO is a simple, straightforward method of obtaining search results. Designed for end-users, it requires neither specialized training nor a knowledge of retrieval commands. Untrained users can obtain information from a broad range of databases without having to learn multiple log-on procedures and search languages. A simple search strategy is created, uploaded to the selected database, and searched. These search strategies are created as SearchMAESTRO leads a user through a series of menu screens (see Figures 3-2 through 3-5); with the information provided, the system identifies the database that is most likely to include the topical information the user needs.

If selection 1 "We pick the database" is chosen, SearchMAESTRO moves the user to the menu shown as Figure 3-3. If selection 1 "Subject" is chosen from this subsequent menu, SearchMAESTRO presents the menu shown as Figure 3-4. If a selection is made from this subject search menu, such as selection 3 "Computers, Sci/Tech, Medicine," SearchMAESTRO presents another menu, shown as Figure 3-5.

The user continues to be prompted with menus until a reasonable narrow search statement is produced, and suitable search subjects are designated by the user. SearchMAESTRO then chooses the most appropriate database and begins to execute the search strategy.

PRESS	TO SELECT	
1	SearchMAESTRO – I	We pick the database
2	SearchMAESTRO – Ii	You pick the database
3	Database directory	
4	News, Instructions	
H	Help	

-->

FIG. 3-2. INITIAL SEARCHMAESTRO MENU PRESENTED TO USER FOR SEARCH STRATEGY DEVELOPMENT

PRESS	TO SELECT
1	Subject
2	Person
3	Place
4	Organization
5	Government Technical Reports
H	Help

-->

FIG. 3-3. "WE PICK THE DATABASE" SEARCHMAESTRO MENU

SearchMAESTRO has also implemented a scan option. Scanning provides the user with a two-stage process for information retrieval. After the user has completed a series of menus that determine the area of information interest, SearchMAESTRO scans a group of preselected databases that include records pertinent to that information area. If selection 3, "Scan group of databases," from the Scan main menu shown as Figure 3-6 is chosen, SearchMAESTRO presents the same sequence of menus to the user as was provided by the typical SearchMAESTRO database-searching features shown as Figures 3-3 through 3-5.

PRESS TO SELECT

1 Current Events

2 Business, Economics

3 Computers, Sci/Tech, Medicine

4 Law, Trademarks, Patents

5 Social Sciences, Education

6 Art, Literature, Entertainment

7 Religion, Philosophy

H Help

-->

FIG. 3-4. STARTING A SUBJECT SEARCH IN SEARCHMAESTRO

PRESS TO SELECT

1 Agriculture

2 Biology

3 Chemistry

4 Computer, engineering, technology

5 Earth sciences, energy

6 Mathematics, physics

7 Medicine, Allied Health

H Help

-->

FIG. 3-5. STEP TWO OF A SUBJECT SEARCH AFTER SELECTION OF THE GENERAL TOPIC "COMPUTERS, SCI/TECH, MEDICINE" IN SEARCHMAESTRO

HOW TO START A SCAN

From System I ("We Choose the Database"),
under most topics, you can choose to:

1 Search a database
2 See list of databases
3 Scan group of databases
H Help

The Scan option leads to a menu that
enables you to scan several related
databases for your topic, see a list of
databases in the scan, or read an
explanation of the scan.

Press (return) to continue ...->

FIG. 3-6. SEARCHMAESTRO SCAN MAIN MENU

As now offered by the DGIS direct-dial feature provided under the menu selection "communicate," SearchMAESTRO also permits users with more knowledge about the databases most suitable for their information needs to select directly any of the individual databases offered by Bibliographic Retrieval Service (BRS), DIALOG, Systems Development Corporation (SDC), NewsNet, Questel, Pergamon/Infoline, and Vu/text Information Services. The user can then dial directly into one of these services and perform database searching without the use of menus, prompts, or search translation capabilities offered by SearchMAESTRO.

3.1.2 Executing the Search Strategy

SearchMAESTRO can automatically translate the user's questions into a format understood by the selected database, dial the proper database vendor, log-on, access a database, and initiate the search. Ten retrieved references at a time are downloaded without intervention by the searcher. The search is performed with no interactive session required by the user with the online databases. An example of how the results are displayed to the user is shown in Figure 3-7.

```
Accessing Network ..... Connected.
Accessing Database Vendor ..... Completed.
Logging on ..... Completed.
Selecting Database ..... Completed.

Each star equals one line of
retrieved data. This may take
several minutes...
**********************************************
**********************************************
**********************************************
**********************************************
********
Search completed .....

There are 10 item(s) which
satisfy your search phrase.

You may wish to PRINT or CAPTURE this data if possible.

Press (return) to see
your search results... -->
```

FIG. 3-7. RETRIEVAL OF CITATIONS USING SEARCHMAESTRO

The user is then able to display (as well as capture and print) a group of 10 citations. An additional 10 citations, if available, can be reviewed. For an additional charge, the abstracts associated with each retrieved citation can also be displayed and captured.

In the "Scan group of databases" option, the user continues through a number of menus until a reasonably definitive subject area is selected and the scan process can begin. After SearchMAESTRO logs automatically onto the appropriate database(s) and conducts the search in a serial fashion, the user is presented with a report that shows how many abstracts or references were in each of the databases scanned that may be relevant to the user's information need (see Figure 3-8). The user may then select one or more of the databases for continued searching.

3.1.3 Performing Post-Processing on the Results of a Search Session

At present, there are no SearchMAESTRO post-processing routines.

Scan completed.....
Press (return) to see
your results ... ->

Computer research and technology profile for: DIGIT/ AND IMAG/

PRESS	TO SELECT	Occurrences	Data Type
1	Compendex	4777	abstracts
2	Computer Database	1017	abstracts
3	Conference Papers Index	436	references
4	INSPEC (1977-to date)	8470	abstracts
5	Microcomputer Index	79	abstracts
6	NTIS	3195	abstracts
7	SciSearch (1984-to date)	638	references
8	Supertech	259	abstracts
H	Explanation of databases		
M	Main menu		

FIG. 3-8. SEARCHMAESTRO DATABASE SCAN REPORT

3.2 GATEWAY UTILITIES

Gateway utilities, (as defined in Section 1.3.2), that can be provided include EM, file handling and editing, and the establishment of subdirectories. SearchMAESTRO supports the accounting for database searching charges of registered users. SearchMAESTRO is capable of providing users with invoices and account statements that are charged to the NTIS accounts.

3.3 GATEWAY SUPPORT

3.3.1 Save Our Search

Telebase now offers a free service that enables users to receive online human assistance while they are connected to the EasyNet system. This service is known as Save Our Search (SOS). The service provides users with access to a bank of search experts. One of these experts can be called upon at any time during a SearchMAESTRO search session. The user need only type an SOS at any prompt to receive this online assistance. DTIC will provide SOS services for DROLS in the future (see Section 7.2.1).

3.3.2 Online Help

Virtually every SearchMAESTRO menu has a help prompt (see Figures 3-2 through 3-5) that can provide guidance and suggestions. As an example, if a user is not sure how to construct a subject search statement using Boolean connectors and requests help, SearchMAESTRO provides the following information:

"CONNECTING WORDS
Don't use small words like: by, from, in, of, the, at.
EX: Joan Arc instead of Joan of Arc

WILD LETTERS
Use / as a "wild letter" at the end of a word.
EX: democ/ will retrieve democracy, democratic, Democrats
Tax/ will retrieve tax, taxes, taxation

LOGIC WORDS (and, or, not)
Use AND to find items common to two or more subjects.
EX: dog AND leash
police AND civilian control
debt AND management

Use OR to find items on either or both subjects.
EX: Bach OR Handel
dog OR cat OR pet

Use NOT to exclude a subject from another.
EX: candy NOT taffy
housing NOT mobile homes

Use ( ) around groups.
EX: (dog OR cat OR pet) AND leash
eskimo/ and (lawyer/ or attorn/)"

With these context-sensitive help messages, SearchMAESTRO provides the end-user with concrete and straightforward assistance.

3.4 GATEWAY FUNCTIONALITY

3.4.1 Menus

With SearchMAESTRO, the user must move straight through the menu structures, upward or downward – never laterally.

3.4.2 Command Mode

No commands per se are available for use with SearchMAESTRO. The user must conduct the search session by means of the menus.

3.4.3 Operating System and Database Management System

SearchMAESTRO is supported by Telebase resources that comprise computers and a telecommunications bank that eases access to many vendors and their databases. These resources are proprietary to Telebase but can be modified and made available for a given client's environment.

3.4.4 Multiple Operations

Generally, SearchMAESTRO performs single operations only. But the scanning feature does permit a user to review the results of a search strategy that has been submitted to many databases without having to provide each database of interest with its own search strategy.

3.4.5 Support of Various Terminal Types

SearchMAESTRO supports a large number of terminal types for access to its prototype system. Virtually any terminal or PC with telecommunications capability can access the DGIS.

3.4.6 Telecommunications Access

Users can gain direct access to the core EasyNet gateway software by dialing directly into the system in Pennsylvania, or by using the DGIS gateway.

3.5 GATEWAY SUMMARY

Clearly, an experienced database searcher would feel hampered by the menu structures and limitations imposed by SearchMAESTRO. Users who know the command language and structure of a given database will perform a more precise search without the assistance of a SearchMAESTRO type of gateway. For example, SearchMAESTRO cannot support the creation of complex searches with sophisticated Boolean strategy statements.

Conversely, the concept of providing the inexperienced user with an easy-to-use, structured approach to database searching is a suitable extension of capabilities offered via DGIS. If users understand that this gateway is designed to provide easy access to one or more databases and to yield results that may not be comprehensive but are designed to point the user toward some pertinent citations from which to

broaden continued searching, SearchMAESTRO has been placed in the proper context.

SECTION 4

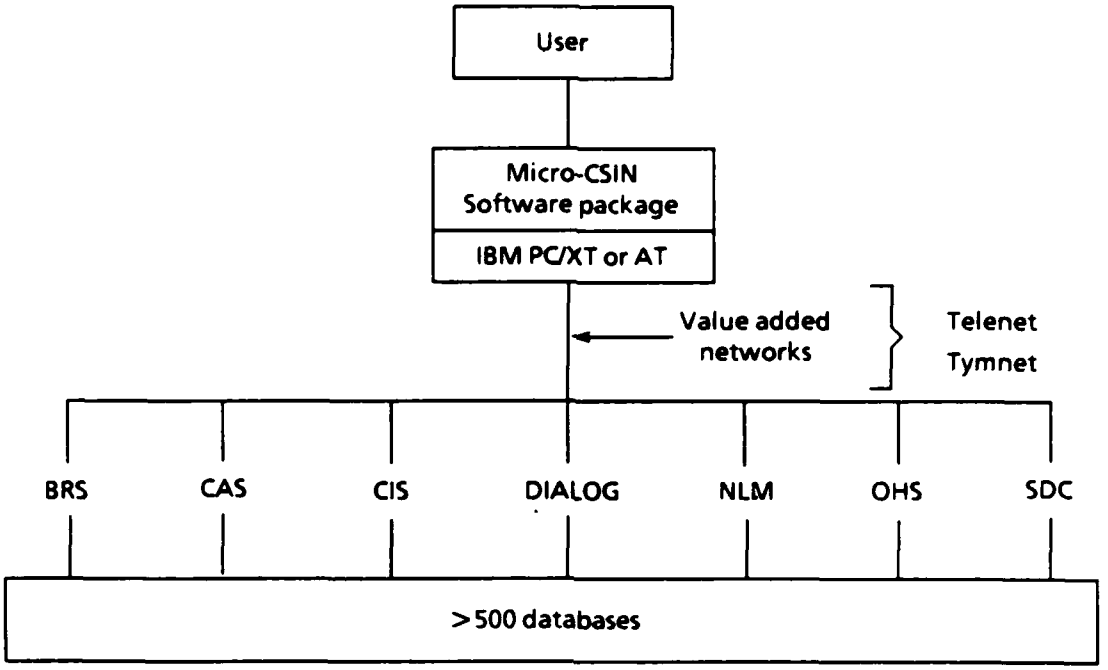
MICRO-CSIN

The Chemical Substances Information Network (CSIN) program was originally mandated by the information requirements of the Toxic Substances Control Act of 1976. Although CSIN was designed at the outset for use with chemical information, later programs, such as Micro-CSIN, have evolved into more generic search tools.

The prototype of CSIN ran on a VAX computer. As the processing power of microcomputers increased, it became possible to implement the CSIN prototype's capabilities on a 68000-based microcomputer that was running the UNIX operating system. Micro-CSIN functionality was adapted to the PC environment, with RS/1 as the foundation. RS/1, a proprietary software package written in the C language, runs on various mainframes, minicomputers, and microcomputers.

The original design and structure of the Micro-CSIN workstation was the result of a cooperative project between the NLM, the Environmental Protection Agency, and the Council on Environmental Quality. The Micro-CSIN workstation was designed to translate a user's request for bibliographic, factual/numeric, and/or chemical identification information into the proper form for interaction with a large number of public sector and commercial databases. Once the connection is established, Micro-CSIN completes the information search request and retrieves the selected information with a minimum of user interaction. A review of the Micro-CSIN workstation concept can be seen in Figure 4-1. The system is designed to provide a consistent, easy-to-use interface that minimizes the amount of information a user needs to interact successfully with the various databases.

Four main types of activities are possible in Micro-CSIN: (1) connect directly to the available vendor systems, including DIALOG; (2) perform file operations; (3) perform script searching; and (4) change the Micro-CSIN settings. The main menu of Micro-CSIN, which allows the user to select from among these functions, appears in Figure 4-2. A schematic view of the Micro-CSIN main menu is provided in Figure 4-3.



Notes: IBM = International Business Machines Corporation; BRS = Bibliographic Retrieval Service; CAS = Chemical Abstracts Service; CIS = (Fein-Marquart's) Chemical Information System, DIALOG = DIALOG Information Service; NLM = National Library of Medicine; OHS = Occupational Health Services; SDC = Systems Development Corporation.

FIG. 4-1. *MICRO-CSIN WORKSTATION CONCEPT*

Main Menu

CHOICES	DESCRIPTION
1 DIRECT	go to direct connection menu
2 FILE	go to file operations
3 PERFORM	a script
4 QUIT	leave Micro-CSIN
5 SETUP	go to setup preferences menu

FIG. 4-2. MICRO-CSIN MAIN MENU

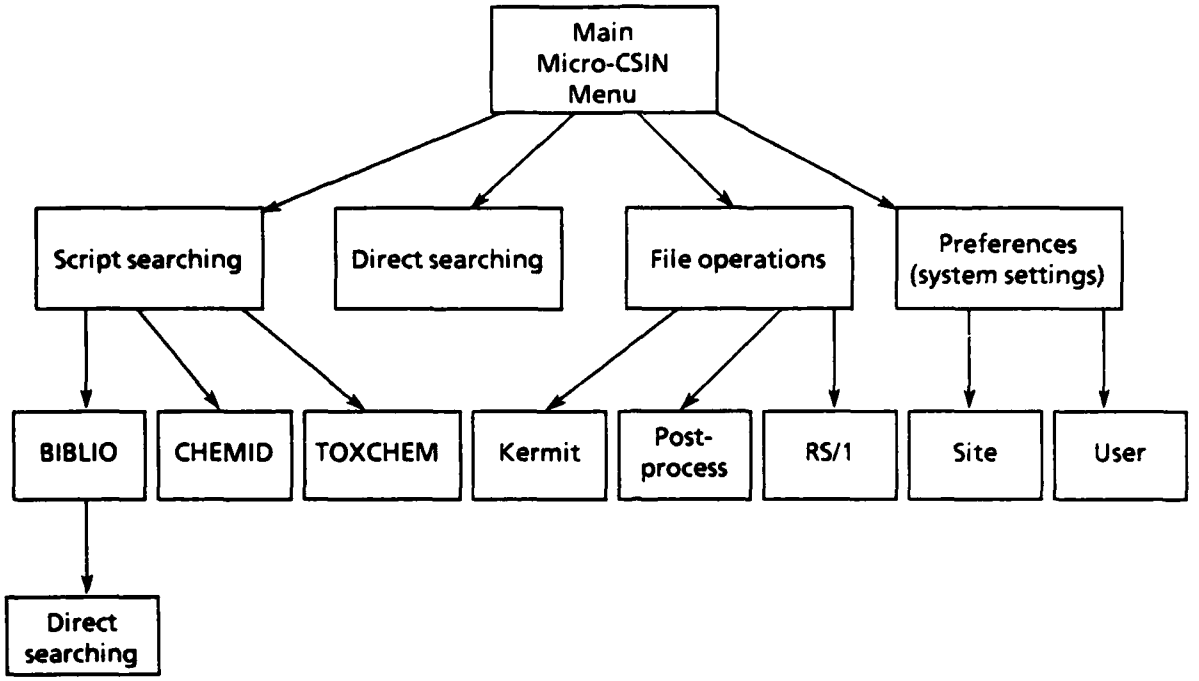


FIG. 4-3. MICRO-CSIN MAIN MENU SCHEMATIC

4.1 GATEWAY FEATURES

4.1.1 Conducting Presearch Activities and Selecting Databases

Micro-CSIN operates in two modes: script and direct. In script mode, Micro-CSIN helps the user phrase an information query; direct mode permits access to a selected database without presearch (script) assistance. Direct mode is similar to the unassisted dial capability of DGIS. For both modes, Micro-CSIN creates a truncated or root word for the subject terms entered by the user and adapts its truncation symbol to the database(s) to be searched.

The definition of a "script," therefore, is an interactive process that is designed to present a step-by-step program to allow the user to setup a search strategy and then identify and retrieve appropriate information.

Three scripts are available from Micro-CSIN: Chemical Identification (CHEMID), Toxic Chemical (TOXCHEM), and Bibliographic (BIBLIO). To request these scripts, the user requests the "PERFORM" option from the Micro-CSIN main menu. All Micro-CSIN scripts have features in common.

Specifically, scripts:

1. Help develop the search strategy
2. Dial the selected system(s)
3. Log onto the system(s)
4. Select the designated database(s)
5. Send search commands in the proper language and format for the system(s)
6. Retrieve specified results
7. Format results automatically according to user specifications in the TOXCHEM script
8. Log out of the system(s)
9. Terminate the connection(s).

For this section of the report, the first four of these capabilities are identified as the presearch activities and database-selection features.

4.1.1.1 CHEMID Script

The CHEMID script searches chemical dictionary files to retrieve chemical identification information, including chemical names, synonyms, trade names, and Chemical Abstracts Service (CAS) Registry Numbers. Databases available for searching include: Chemline from NLM; Chemdex from SDC; Structure and Nomenclature Search System (SANSS) from Chemical Information System (CIS); Chemsis, Chemzero, and Chemname from DIALOG; and Registry File from CAS. Once the CHEMID script has retrieved the needed data, the results can be transformed into a standard format. The resulting file can be further used in the TOXCHEM or BIBLIO scripts and Direct Mode searching.

A user requests the CHEMID script from the Micro-CSIN main menu selection of "PERFORM". When the CHEMID search choice has been made, the following menu (Figure 4-4) is presented:

CHEMID Search Choices

CHOICES	DESCRIPTION
1 REGNUM	search by chemical registry number
2 NAME	search based on chemical name
3 PERFORM	a script
4 QUIT	leave Micro-CSIN
5 SETUP	go to setup preferences menu

FIG. 4-4. CHEMID SCRIPT SETUP MENU

Under either search strategy option, multiple chemical names or registry numbers can be entered until the full set has been defined. If an incorrect CAS Registry Number is entered, Micro-CSIN presents an error message.

Once the user has completed the script choices, Micro-CSIN asks which databases are to be searched. Each of the available databases contains information about preferred chemical names, synonyms, CAS Registry Numbers, and molecular formulas. Usually only one database is needed for successful completion of the search strategy.

4.1.1.2 TOXCHEM Script

The TOXCHEM script searches NLM's Toxicology Network (TOXNET) and Hazardous Substances Data Bank (HSDB), CIS's Oil and Hazardous Materials Technical Assistance Data System (OHMTADS), and the Occupational Health Services' (OHS') Hazardline. This script is designed to retrieve the following factual/ numeric categories of chemical information: chemical identification, chemical/ physical properties, critical exposure levels, emergency response/handling, toxicity, production and use, environmental effects/concentrations, laboratory and

monitoring methods, and regulations. Unlike the CHEMID script, TOXCHEM is designed to search on only one chemical at a time.

Setting up a search with the TOXCHEM script requires the entry of three items by the user: chemical identifier (e.g., a name, CAS Registry Number, or predefined file of identifiers already established in conjunction with the TOXCHEM script), the name of the profile to be used in the search, and the name of the output report name in which to store the search results. The information retrieval profile, the second item required as user input, determines which databases will be searched, which fields will be retrieved, and how the report is to be organized. Users can store as many profiles as they wish on topics of information they know to be of recurring interest. If the user is unsure about which profiles have been stored or are available as public profiles, the entire list of available profiles can be displayed.

Building a TOXCHEM profile can require a substantial amount of a user's time. However, the process has been designed as a menu-driven activity, with online assistance and default fields already selected to expedite the process. The setup of a TOXCHEM profile begins with the following presentation (Figure 4-5) to the user:

Topics of Information

Choice		Description	Status
1	ID	Identification Information	[NO]
2	CHEM/PHYS	Chemical/Physical Properties	[NO]
3	CRIT-EXP	Critical Exposure Levels	[NO]
4	EMERG	Emergency Response/Handling Data	[NO]
5	TOX	Toxicity Information	[NO]
6	PROD/USE	Production and Use Information	[NO]
7	ENVIRON	Environmental Fate/Concentrations	[NO]
8	LAB/MON	Laboratory and Monitoring Methods	[NO]
9	REGS	Regulations	[NO]

FIG. 4-5. TOXCHEM PROFILE SETUP

The Status Line indicates the selections the user has made for the fields within each topic that will be searched as the profile is sent to a selected database. "NO" means no fields have been chosen. Other available choices indicate that all fields, a default list, or a partial list is to be selected. Once a profile is completed, it is saved for recall when needed.

4.1.1.3 BIBLIO Script

The general bibliographic script, BIBLIO, permits searches of approximately 230 bibliographic databases in a wide range of subject areas from any of the vendor systems the Micro-CSIN workstation accesses – CAS, BRS, DIALOG, NLM, and SDC.

If the selection PERFORM is requested from the Micro-CSIN main menu, the system first displays the BIBLIO menu, as follows:

CHOICES	DESCRIPTION
1 SEARCH	edit or create a search strategy
2 DATABASE	edit or create a database list
3 RUN	execute a search

FIG. 4-6. MICRO-CSIN GENERIC BIBLIOGRAPHIC SCRIPT

If the user specifies editing or creation of a search strategy (Choice 1), Micro-CSIN asks whether the user wishes to create a new strategy. If the answer is "NO," the system asks the user for the name of the stored strategy. Once named, the file is displayed, and the user has the option of editing the strategy. If the user wishes to create a new search strategy, Micro-CSIN presents a search template (Figure 4-7) that is to be filled in.

A search strategy can include any combination of template categories. A user may enter either ad hoc search terms or predefined keyword lists containing search terms. These keyword lists are first created as a separate function by the user and then stored in a file. These previously developed keyword lists can then be called up and brought into the search strategy. A mix of ad hoc keywords and predefined keyword lists can be used in the same strategy.

Micro-CSIN links the individual categories automatically so that the sequence of searching is: authors AND chemicals AND keywords AND years AND language. Within the individual categories of author, chemical, and keyword, any combination of three basic logical operations (AND, OR, AND NOT) can be used. A completed Search Strategy Menu is illustrated in Figure 4-8. When the user has created a new

Search Strategy Template

Categories are ANDed together

CHOICES	DESCRIPTION
1 AUTHOR	
2 CHEMICAL	
3 KEYWORDS	
4 YEAR LIMIT	NO YEAR LIMITATION
5 LANGUAGE	NO LANGUAGE LIMITATION
6 FILE	containing search strategy

Not Connected

Menu choice:

FIG. 4-7. MICRO-CSIN SEARCH STRATEGY TEMPLATE

Search Strategy Template

Categories are ANDed together

CHOICES	DESCRIPTION
1 AUTHOR	burnside, j. AND craig, p. AND guthrie, g.
2 CHEMICAL	(butylene oxide OR xylene.id) AND NOT ddt
3 KEYWORDS	(sort(2)key# OR systematic) AND algorithm key
4 YEAR LIMIT	1982 - 1984
5 LANGUAGE	ENGLISH
6 FILE	

FIG. 4-8. MICRO-CSIN COMPLETED SEARCH STRATEGY MENU

strategy or has modified an existing one, Micro-CSIN displays the File Write Menu (Figure 4-9).

File Write Menu

CHOICES	DESCRIPTION
1 SAVE	the new or modified object
2 WRITE	a modified object to a new file
3 CONT	continue editing the current object
4 QUIT	abandon the modified project

FIG. 4-9. FILE WRITE MENU

In addition to developing a search strategy before going to a database, the user needs to indicate which databases are most appropriate for searches. If the user chooses "DATABASE" from the generic BIBLIO script (see Figure 4-6), the user is asked whether to use a new database list or an existing one. If an existing list is chosen, the user can add or delete databases from the list. If a new list is to be compiled, the user is presented with the Database Subject Area menu (see Figure 4-10). A user can select from one of these broad subject areas and then add or delete databases from a given list from the Subject Area menu. For example, if the user selects "1: AEROSPACE," Micro-CSIN displays a menu of Aerospace Databases, as in Figure 4-11.

The user can return to any broad subject area and select additional databases. When all databases have been selected, Micro-CSIN shows the user which ones have been chosen and the sequence in which they will be searched. The user has the option of reordering this sequence.

The process of saving a database list is identical with saving a search strategy. When the user has instructed Micro-CSIN to save the strategy, a File Write menu is displayed, enabling the user to "SAVE," "WRITE," "CONT," or "QUIT."

4.1.1.4 Direct Mode

The direct mode (selection of "DIRECT" from the main menu) provides the user with broader flexibility in accessing and performing commands in a given system

Select Databases by Subject Area

	CHOICES	DESCRIPTION		CHOICES	DESCRIPTION
1	AEROSPACE	Aerospace Data	10	GEOLOGY	Geological Data
2	AGRICULTURE	Agriculture Data	11	HUMANITIES	Humanities Data
3	BUSINESS	Business Data	12	HYDROLOGY	Hydrological Data
4	CHEMISTRY	Chemical Data	13	MEDICINE	Medical Data
5	COMPUTER	Computer Data	14	PATENTS	Patent Data
6	ENERGY	Energy Data	15	POLITICAL	Politics, Current Ev
7	ENGINEER	Engineering Data	16	REGULATIONS	Government Regs
8	ENVIRONM	Environmental Data	17	SOCIAL	Social Sciences
9	GENERAL	General Interest	18	TOXICOLO	Toxicology Data

0 databases selected

FIG. 4-10. MICRO-CSIN DATABASE SUBJECT AREA MENU FOR GENERIC BIBLIO SCRIPT

AEROSPACE (Select Databases by Number)

	DATABASE	VENDOR	
1	COMPENDEX	(BRS)	[ ]
2	NTIS	(BRS)	[ ]
3	AEROSPACE DATABASE	(DIA)	[ ]
4	COMPENDEX	(DIA)	[ ]
5	METRL & GEOASTROPHYS	(DIA)	[ ]
6	COMPENDEX	(SDC)	[ ]
7	EI ENGINEERING MTGS	(SDC)	[ ]
8	NTIS	(SDC)	[ ]

FIG. 4-11. MICRO-CSIN MENU OF AEROSPACE DATABASES

and one of its databases. This feature is similar to the "Unassisted Dial into Other Systems" available on the DGIS. In direct mode, Micro-CSIN can connect the user to any of the vendor systems available in the script mode plus DIALOG. However, the user must be quite familiar with the commands, syntax, and features of each accessed vendor system, since there is no interface – usually provided by such a feature as a TOXCHEM or BIBLIO script – between the user and the database.

If the user wishes to employ the direct connection mode, the selection must first be made from the main menu (see Figure 4-2). After the menu choice "DIRECT" is made, the user is presented with the menu shown as Figure 4-12.

Direct Connection Menu

CHOICES		DESCRIPTION
1	BEGIN	data capture
2	CONNECT	to a component system
3	DISCONNECT	from a component system
4	END	data capture
5	RESUME	component connection
6	SUBSTITUTE	a file or keyword list
7	INTERACT	run an INTERACT script

FIG. 4-12. DIRECT CONNECTION MENU

To connect to the component system, the user selects "2: CONNECT" from the Direct Connection menu. Micro-CSIN then displays the Component System Choices menu as shown in Figure 4-13. The user can select any of the available systems on the Component System Choices menu. (BRS will be added in the near future.) Micro-CSIN dials the system automatically and completes the log-on. From this point on, the user is connected directly to the system specified, until Micro-CSIN is asked to "DISCONNECT."

Once a user is connected to a component system, the data connection can be put "on hold," and the user can return to Micro-CSIN to have it provide search assistance. This assistance can occur in several ways, for example, by use of a

Component System Choices

CHOICES	DESCRIPTION
1 CIS	NIH/EPA Chemical Information Systems
2 NLM	National Library of Medicine
3 OHS	Occupational Health Services, Inc.
4 SDC	Systems Development Corporation ORBIT
5 DIA	Dialog Information Systems
6 TOX	TOXNET – National Library of Medicine
7 BRS	BRS is not available
8 CAS	Chemical Abstracts Service

Note: EPA = Environmental Protection Agency.

FIG. 4-13. MICRO-CSIN SYSTEM CHOICES MENU

keyword file stored in Micro-CSIN that can be introduced to a database for retrieval needs.

Once search results begin to appear, the user can begin downloading the citations by requesting the "BEGIN" option. This capability also captures all interactions with the vendor systems.

4.1.2 Executing the Search Strategy

4.1.2.1 CHEMID Script

Once the CHEMID script has been completed and the database selected, Micro-CSIN automatically dials the system, performs the log-on sequences, searches for the information, and retrieves and stores the results in the predesignated file. During the search process, if telecommunications problems are encountered, they are reported to the user. Once the search has been completed, the user may run the CHEMID script again in another database.

4.1.2.2 TOXCHEM Script

A TOXCHEM script profile that has been saved (see Section 4.1.1.2) can be run against one or more of the selected databases. These searches are designed to be conducted quickly, to present emergency responses about given chemicals.

4.1.2.3 BIBLIO Script

When the user has constructed a BIBLIO search strategy and chosen a database list, the search can be conducted. From the top level of the BIBLIO menu, the user selects the "RUN" command and Micro-CSIN prompts for the file names of the search strategy and database list. Micro-CSIN then moves on to logging-on the first selected database. Information on the activities that are occurring while connected to the selected database(s) is displayed on the screen's status line. The search continues through each of the requested database(s).

4.1.3 Performing Post-Processing on the Results of a Search Session

4.1.3.1 CHEMID Script

Micro-CSIN transforms the results of a *CHEMID* search automatically, no matter which database(s) is/are used, and stores them in a standard CHEMID format. These converted CHEMID files can be used further for the "CHEMICAL" line of the BIBLIO search template.

4.1.3.2 TOXCHEM Script

The purpose of a TOXCHEM search is to provide the user with factual or numeric information on an individual chemical substance. Consequently, there is limited post-processing activity, the merger of retrieval results from multiple sources is not needed. However, TOXCHEM does rearrange the results into a standard format, irrespective of the database that was searched, and presents the results in a topical report.

4.1.3.3 BIBLIO Script

After each database has been searched, Micro-CSIN displays a summary of the search results for each search template category (author, chemical, keyword, etc.).

As an example, the user may be provided with the actual number of retrieved citations, as shown in Figure 4-14.

Select a database to Print

(Press return to go to next database)

	Description	
1	After author search	53 postings
2	After chemical search	12 postings
3	After keyword search	8 postings
4	After year limits	5 postings
5	After language limit	5 postings

FIG. 4-14. MICRO-CSIN DISPLAY OF RETRIEVED CITATIONS USING BIBLIO SCRIPT

By entering a <Return>, the user can move on to the display for the next database that was searched.

To print the results of a search session, the user has the choice of five print options and can select a range of records to be printed.

With records retrieved from a BIBLIO search strategy, further post-processing is possible. Like DGIS, Micro-CSIN has a post-processing menu that offers the following capabilities: transform records to a standard bibliographic format, merge the transformed files, and sort a transformed file.

4.2 GATEWAY UTILITIES

4.2.1 Electronic Mail

No EM feature is available on the Micro-CSIN software, but public and private files can be established under Micro-CSIN. One user may leave a keyword list(s) or recommended search strategy in a public file for other users to review and use. External EM systems can be used by access via the Micro-CSIN direct mode feature.

4.2.2 File Handling and Editing Capabilities

Every activity that takes place in Micro-CSIN creates a file that is subsequently stored on the microcomputer. Micro-CSIN labels each file with an extension name, based on which operation – e.g., the CHEMID script – generated the file. These files may be copied, deleted, edited, listed, printed, and renamed. Two types of files can be supported by Micro-CSIN – public and private. Each file type is stored in a different subdirectory on the system. Users can store private files separately from files that they may wish to share with other Micro-CSIN users who use the same workstation.

4.3 GATEWAY SUPPORT

4.3.1 Online Help

Three types of help assistance are provided by Micro-CSIN. A user who types "help" in response to a menu prompt receives general information on how to switch menus and obtain more detailed help information. At any point in Micro-CSIN, if a "?" is keyed, more specific help is provided. Finally, if Micro-CSIN has prompted the user for a file name and the user is unsure of the names of the files that have been established – e.g., a BIBLIO script – a "?" produces the names of the available files.

4.3.2 Users Directory

Since this package is a PC-based single-user gateway, there is no recording of Micro-CSIN users in a centralized file.

4.4 GATEWAY FUNCTIONALITY

4.4.1 Menus

Micro-CSIN's menus are linked hierarchically in sequences. Each selection made from a menu leads the user to the next menu in the sequence or to the next prompt for the current operation. A user cannot jump to a different path and must move up or down the currently selected path completely in order to start on a new

process. All Micro-CSIN menus are displayed on the top portion of the microcomputer's screen, and all have the same general appearance:

- Title, e.g., "Search Strategy Template"
- List of one-word choices available on the menu, e.g., "KEYWORDS"
- Brief description of what each choice provides, e.g. "containing search strategy"
- Number preceding each choice.

As an example, see Figure 4-7, "Micro-CSIN Search Strategy Template."

Menu choices can be made in one of three ways: enter the number of the menu option, enter the name of the option or its unique prefix, e.g., "f" for File Operations, or move the cursor to the menu option.

Also present on a screen is a highlighted status line (see Figure 4-7). This line tells the user whether Micro-CSIN is connected to a vendor system. The line also displays the status of an activity that the user has requested of Micro-CSIN – e.g., conduct a TOXCHEM script search or rename a file.

4.4.2 Command Mode

The ability to enter abbreviated selections can be used throughout Micro-CSIN. Menu selections can usually be shortened to the first letter, e.g., "p" for "perform." In some instances, a user may have to enter two or three characters to make the selection unambiguous. If a nonunique selection is made, Micro-CSIN responds, like the DGIS COPS feature, with the possibilities and asks the user to select one.

4.4.3 Operating System and Database Management System

The most recent version of the Micro-CSIN code has been developed under C and is designed to operate on an International Business Machines Corporation (IBM)-PC/XT or AT, running Disk Operating System (DOS), and equipped with 512K random access memory (RAM) and at least a hard disk of 10-megabyte capacity. The system must also have a floppy disk drive, a printer, a monitor, and appropriate cabling. A 1200-baud modem (auto answer/auto dial) and an IBM asynchronous communications expansion card are also required. At present, the

Hayes Smartmodem, Vadic 3451PA, and user-dialed modem are supported for dialing out to selected databases.

The evolution of the technology that has supported CSIN and Micro-CSIN is summarized in Figure 4-15.

4.4.4 Support of Various Terminal Types

Although there is a multi-user version of CSIN available, the most recent version with the features described in this section is available as a PC-DOS based package only.

4.4.5 Telecommunications Access

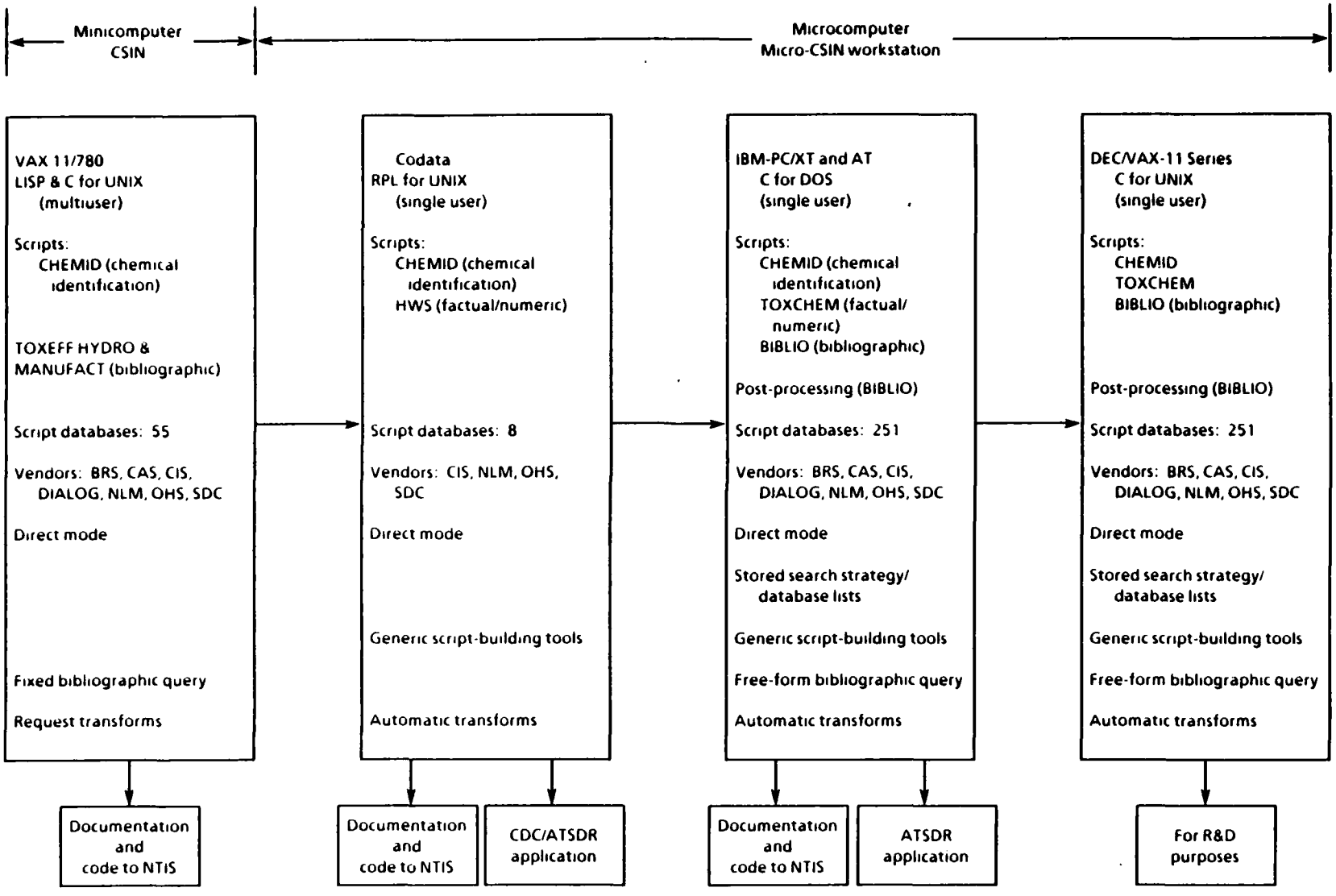
The Micro-CSIN package communicates with external systems and databases through a number of networks. Connections are established through the modem attached to the workstation/microcomputer's communications port. Micro-CSIN can connect a user to a database by either a direct connection or a public connection network, such as Telenet.

Micro-CSIN also includes the *file transfer and communications* program called Kermit, which was developed at Columbia University. While in Micro-CSIN, the user can employ Kermit to transfer files to or from any other computer with which this program has an established protocol, whether the machine to be tapped is a mainframe, minicomputer, or microcomputer.

4.5 GATEWAY SUMMARY

Micro-CSIN, although developed originally for searching chemical information, has several features not offered by the other three gateway products that are available in its more generic offerings. It provides script searching (search support) for assisting a user in establishing an efficient search strategy for capturing chemical identification information and factual/numeric chemical data that can be downloaded from as many as eight vendor systems. The results of such searching can be integrated into a single report, organized by subject, with features comparable to those of DGIS.

Micro-CSIN is the only gateway that offers the ability to leave and return to this script mode. Micro-CSIN also comes with more than 20 keyword lists. These are lists of search terms in a wide variety of subject areas that have been compiled by



Notes: ATSDR = Agency for Toxic Substances and Disease Registry, CDC = Center for Disease Control, LISP = LISt Processing

FIG. 4-15. EVOLUTION OF MICRO-CSIN TECHNOLOGY

expert searchers, taking into account specific database and vendor system features. These lists can be used to help expedite a database search. Micro-CSIN also permits predesignation of databases that appear to be most pertinent for the user to send completed search scripts to selected databases for searching and retrieval.

As Micro-CSIN evolves (see Section 7.3), it will continue to expand, eventually becoming the foundation for a gateway to access the whole realm of biomedical sciences.

SECTION 5

GRATEFUL MED

Grateful MED is a PC-based, NLM-developed software package that enables searchers to use their PCs to search NLM's files – specifically MEDLARS OnLINE (MEDLINE), MEDLINE back files, and CATaloging OnLINE (CATLINE). The computer files of NLM contain more than 20 years' worth of bibliographic data from over 3,000 major medical journals in the MEDLINE file and a book catalog (CATLINE file). In total, a Grateful MED search can access more than 6 million journal articles or 600,000 book titles. Like SearchMAESTRO, it is a simple design that allows an end-user to compose a straightforward search strategy and quickly receive retrieved citations.

5.1 GATEWAY FEATURES

Basically, Grateful MED operates in four steps: (1) formulating the search, (2) accessing the NLM computer in Bethesda, Md., (3) executing the search and disconnecting from the NLM computer, and (4) displaying search results and, optionally, printing the results. The main menu for Grateful MED is shown in Figure 5-1.

SELECT THE ACTION YOU WANT TO TAKE

1. Search recent journals (MEDLINE) using the INPUT form screen.
2. Search books (CATLINE) using the INPUT form screen.
3. Search directly without using the INPUT form screen.
4. Review the results of last search again.
5. See the GRATEFUL MED system overview description.
6. Run SETUP to enter new telephone numbers/log-on codes.
7. Quit.

ENTER TOP ROW NUMBER OR FUNCTION KEY FOR YOUR CHOICE.

FIG. 5-1. GRATEFUL MED MAIN MENU

5.1.1 Conducting Presearch Activities and Selecting Databases

Figure 5-2 shows the screen that is first presented when the user begins a typical Grateful MED search in the MEDLINE database (option 1 on the main menu).

INPUT YOUR SEARCH

FILL IN THE APPLICABLE LINES, E.G., TYPE AUTHOR'S NAME IN THE AUTHOR LINE. HIT ENTER TO GO TO NEXT LINE. FOR HELP HIT HOME. TO START OVER HIT ESC.

AUTHOR/NAME
TITLE WORDS
SUBJECT WORDS
  2ND SUBJECT
  3RD SUBJECT
  4TH SUBJECT
ENGLISH ONLY
REVIEW ONLY
JOURNAL ABBREV

FIG. 5-2. GRATEFUL MED SEARCH INPUT SCREEN

We shall describe the process for completing this INPUT screen in a line-by-line sequence. It is not necessary for the user to complete every line of the INPUT screen in order for a search to be accepted on the NLM computer.

5.1.1.1 Author/Name

An author's name is input by the user as first name followed by initial and surname, e.g., "JOHN E. SMITH". Grateful MED is able to change the name to a recognizable pattern for MEDLINE or CATLINE searching. Once the user has entered the author's name, Grateful MED transforms and replaces the author information with the form that will be accepted by the NLM databases. After transformation, therefore, JOHN E. SMITH appears as "SMITH JE:" and "SMITH, JOHN E:" on the AUTHOR/NAME line.

5.1.1.2 Title Words

Significant terms that the user thinks essential to a given search can be input on the "TITLE WORDS" line. These words are ANDed together, and all must appear in the title, for the NLM database record to be retrieved.

5.1.1.3 Subject Words

The words entered on the four subject lines of the search form are searched as both text words from the title and abstract and Medical Subject Headings (MeSH) terms. On each separate subject line, only one main word or group of synonyms can be entered. These subject words are ANDed together. For the record to be selected, every word must appear in the MEDLINE citation. Adding a ":" to a core word (such as "drown", changing it to "drown:") enables the user to search for such words as "drowned" and "drowning", words that have any number of characters in addition to the core word. Adding a "#" to a core word ("electrolyte" making it "electrolyte#") enables the user to retrieve a word that has only one character in addition to the core word ("electrolytes").

Synonyms can be entered together on the same subject line. These words will be ORed together, in that only one of the group need appear in the MEDLINE citation for the record to be retrieved. If the user wishes the synonyms to be ANDed together, the group of synonyms should be preceded by the word "ALL".

MeSH terms, which make up NLM's controlled vocabulary or thesaurus, can be entered on a subject line if preceded by a "/" – e.g., "/pulmonary emphysema". MeSH headings are used by professional indexers to describe a journal article for the MEDLINE database. This symbol signals Grateful MED to search for records using the MeSH heading(s). Users who are more familiar with MEDLINE searching can also employ the "EXPLODE" command. The EXPLODE command requests that a "tree" or logical grouping of MeSH terms be used to retrieve database citations. Some of these "trees" have been predefined and stored in the NLM computer and can be called up for searching. For example, "/EXPLODE VASOPRESSINS" will cause Grateful MED to ask for the terms "VASOPRESSINS" as well as the narrower terms "ARGIPRESSIN," "DESMOPRESSIN," "LYPRESSIN," "FELYPRESSIN," and "ORNIPRESSIN."

5.1.1.4 English Only

If the user wishes to see citations from English-language publications only, the designation can be made in the "ENGLISH ONLY" line.

5.1.1.5 Review Only

An approach similar to that for "ENGLISH ONLY" is employed if the user wishes to see review articles only. The designation is made on the "REVIEW ONLY" line of the search strategy template screen.

5.1.1.6 Journal Abbreviation

Journal title abbreviations can also be used to help narrow or specify the MEDLINE search. These abbreviations must be entered in the Index Medicus format. If the user is unsure of an abbreviation, Grateful MED provides access to more than 100 abbreviations of English-language clinical journals, which the user can review by pressing the HOME key when on this search template line.

5.1.2 Executing the Search Strategy

Once the INPUT form screen has been completed, Grateful MED first combines words on each subject line with "OR" and then combines all the completed subject lines with "AND." Individual subject words are searched as text words in the abstract or title and as single-word MeSH headings. When the search strategy has been completed and the user wishes to submit the request to the NLM database(s), the following two questions are asked:

"DO YOU WANT TO RETRIEVE ABSTRACTS (Y/N)?"

The user has the option of retrieving abstracts with MEDLINE search only. Abstracts are not available on the CATLINE database.

"OK TO GO ON TO SEARCH (Y/N)?"

This question provides the user with a final chance to review and make any changes on the INPUT screen before the search is submitted to the NLM database. If the user indicates 'no', Grateful MED allows changes to be made to any INPUT screen by means of the cursor positioning keys. If the user answers 'yes', Grateful MED begins the telecommunications process to access the NLM

computer. Once the process has begun, the user cannot change the search strategy. The message displayed to the user is as follows:

```
I AM NOW GOING TO CALL THE NLM COMPUTER
AND RUN YOUR SEARCH. DON'T WORRY ABOUT
READING THE RESULTS AS THEY ARE RECEIVED;
I'LL SHOW THEM TO YOU AGAIN LATER.
DON'T TOUCH THE KEYBOARD UNLESS I ASK A
QUESTION OR SOMETHING GOES WRONG.
IF SOMETHING DOES GO WRONG,
HIT THE ESC KEY SEVERAL TIMES.
IF THIS DOESN'T WORK, HIT THE CTRL-BREAK KEYS
AND START OVER.

RESETTING MODEM
```

Grateful MED dials the NLM computer, establishes a connection, enters the user's ID/password and logs into the system. Then the desired file is selected, and the search/retrieval interaction begins.

While Grateful MED is interacting with the NLM system and begins displaying the retrieved references, the user can stop the search by pressing the spacebar. Grateful MED then logs off the NLM system. By continuing to employ Grateful MED features, the user has offline access to retrieved references that have been displayed until the time of log-off.

Grateful MED can also connect a user to the NLM computer and allow for direct interaction with MEDLARS without use of the INPUT screen prompts. The user does this by requesting option 3 from the main menu.

5.1.3 Performing Post-Processing on the Results of a Search Session

At the end of the search session, Grateful MED displays the following screen:

```
YOUR SEARCH FOUND _____ REFERENCES.
YOU SAVED _____ OF THESE TO REVIEW NOW.
THE COST OF THIS SEARCH WAS $_____.

DO YOU HAVE A PRINTER AND WANT TO PRINT (Y/N)?
```

Before printing or displaying the retrieved citations, Grateful MED reformats each reference into the following components:

- Authors (up to four – with the notation "AND OTHER" if there are more)
- Title (in brackets if not in English)
- Language (only if not in English)
- Source (journal, date, volume, issue, pages)
- Unique identifier (such as the International Standard Serial Number)
- Abstract (if requested on the INPUT form screen and present in the retrieved citation).

Whether the user decides to print the retrieved citations or not, after the search is finished, a user can choose among the retrieved citations. Then, with an optional feature, citations that the user judges to be pertinent can be printed out. If, at any time, the user wishes to print out all the other citations in the retrieved set, another option makes this possible.

Based on the MeSH headings attached to records that the user has judged to be pertinent, Grateful MED can recommend other MeSH headings for a more comprehensive search. The user does not receive MeSH suggestions unless there are MeSH subject terms common to the majority of the citations that were judged pertinent. Based on this pertinency feedback, the user can conduct another search with the suggested MeSH headings.

If no citations are retrieved, Grateful MED displays the following at the end of the search:

```
THE COST OF THE SEARCH WAS $_____

BUT THE SEARCH FOUND NOTHING.
DO YOU WANT TO SEE THE INTERACTION (Y/N)?
(HIT HOME FOR HELP)
```

The user can then be shown the activity between Grateful MED and the NLM database, to determine why no citations were retrieved. The user can look to see if a word was misspelled, if inappropriate or unrecognized terms were used, if terms that

were too general or too narrow were used, or if some terms should be dropped from the search strategy.

Another searching option offered by Grateful MED is the ability to run a search against older MEDLINE files. The following question is posed to the user if after reviewing the results of a MEDLINE file search, a search of older materials seems warranted:

"DO YOU WANT TO RUN THE SAME SEARCH AGAINST OLDER MATERIAL (Y/N)?"

If the user answers 'yes', Grateful MED asks:

"HOW MANY YEARS BACK BEFORE THE CURRENT FILE?" _____

The current MEDLINE file always contains references to materials 2 or 3 years old; the older MEDLINE files go back to 1966.

5.2 GATEWAY UTILITIES

Grateful MED, like Micro-CSIN, is a PC-based software package. As such, it does not support features that are designed for software to support a multi-user population.

5.2.1 Electronic Mail

No EM capability is available with Grateful MED. But Grateful MED can – by use of the direct mode – enable a user to connect to an external system where EM capabilities are available.

5.2.2 File Handling and Editing Capabilities

Files are created whenever a search is conducted and downloaded to Grateful MED for user analysis and review. These downloaded files are named by the package without user intervention. Every time the user conducts a new search, the downloaded file is overwritten. Search results are retained for the most recent search run only, as can be seen from the available options on the main menu: "4. Review the results of last search again."

5.3 GATEWAY SUPPORT

5.3.1 Online Help

If a user who needs help at any time during a search session presses the HOME key on the PC, detailed instructions are immediately displayed in context on the screen. The user is then returned to the place left in Grateful MED.

5.3.2 Users Directory

There is no recording of users because Grateful MED is a single-user package.

5.4 GATEWAY FUNCTIONALITY

5.4.1 Menus

There are specific sequences in which the screens are presented to a user. A user must progress through the relevant series of menus before starting another search.

5.4.2 Command Mode

A user who has become familiar with the menu screens can bypass them. These are the bypasses:

1 or M for MEDLINE INPUT form screen

2 or C for CATLINE INPUT form screen

3 or D for searching directly without using the INPUT form screen

4 or R for reviewing results of the last search

5.4.3 Operating System and Database Management System

Grateful MED can be operated on any member of the IBM PC family and most IBM-compatibles with at least 256K RAM and one 360K double-sided, double-density disk drive. DOS version 2.0 or higher and a Hayes Smartmodem (or other fully compatible modem) are required.

5.4.4 Support of Various Types of Terminals

This gateway operates only on a PC that has the configuration designated in Section 5.4.3.

5.4.5 Telecommunications Access

This gateway operates only on a PC that has the configuration designated in Section 5.4.3. Grateful MED is not designed to be accessed on a centralized computer by multiple users, but only in a one-on-one PC environment.

5.5 GATEWAY SUMMARY

Grateful MED is a straightforward, yet sophisticated, search capability for the professional end-user who requires access to the medical literature. Transformations of a user's search strategy, as entered on the INPUT screen, into an "acceptable" NLM structure removes much of the confusion and frustration when an inexperienced user is trying to formulate a search strategy. The user need not be concerned with a given structure for the search to be run successfully against the MEDLINE or CATLINE databases. The online support for locating MeSH terms and Index Medicus journal abbreviations is a feature of immense benefit to the user, because it provides additional facilitated guidance into the fields in the two databases. The post-processing functions concentrate on providing the user with continuing search guidance by suggesting more pertinent MeSH terms, based on the judgments the user makes regarding the retrieved citations. Presenting the retrieved citations in a predefined standard format also reduces the likelihood that an inexperienced end-user will be overwhelmed by too much data, as would be presented by a full MEDLINE or CATLINE record.

It should be noted that the capabilities provided by Grateful MED were perhaps easier to design and produce, since the package can work with two databases only. The other three gateways reviewed in this project attempt to access a multitude of databases. Each of those databases can represent a unique record structure and have a different underlying controlled vocabulary. However, the accomplishments of Grateful MED should be assessed as the other gateways continue to expand in presearch and post-processing.

SECTION 6

COMPARISON OF GATEWAYS

For the most meaningful comparisons, the gateways have been divided into two groups: DGIS-and-Micro-CSIN and Grateful-MED-and-SearchMAESTRO. The gateways within these two groups are functionally similar, though DGIS and Micro-CSIN are geared to support a more sophisticated user group, typically intermediaries, and SearchMAESTRO and Grateful MED are designed to retrieve database search results quickly with limited end-user interaction and typically serve inexperienced searchers. The two groups represent a mix of hardware environments: DGIS and SearchMAESTRO were developed for placement on mainframe or large minicomputers; Micro-CSIN and Grateful MED were created for the PC environment. One-for-one comparisons between one gateway that operates on a large computer and another that works on a PC will result in an asymmetric presentation. For example, a PC-based package cannot support multiple simultaneous operations, because it is based on DOS.

The gateways are compared here in a series of tables and accompanying text.

6.1 COMPARISON OF GATEWAY FEATURES

6.1.1 DGIS and Micro-CSIN

The distinct differences in features between these two gateways center on the presearch search strategy set-up and the manner of the database searching interaction with which the user is provided. DGIS, given its current capabilities, cannot provide the user with guidance in search strategy or database selection. Micro-CSIN can help a user to structure a search strategy and predefine those databases that may be most suitable for search execution. Micro-CSIN can also submit a search strategy serially to multiple databases without any request from the user. Both packages are versatile in their post-processing capabilities.

TABLE 6-1

GATEWAY FEATURES: DGIS AND MICRO-CSIN

Component	DGIS	Micro-CSIN
Presearch activities and database selection	No presearch or database selection capabilities now offered. Two enhancement projects underway to provide: CCL for presearch support and Directory of Resources for database selection needs.	Three "scripts" available to help user phrase information query. Scripts are step-by-step process for search strategy set-up and can be saved for future use. Databases are chosen by user via menu selection. Database selection lists can be saved and applied to future search strategies.
Search execution	User is actively involved in online interaction with the database during search execution. Search strategies may also be uploaded via a text editor. No intervention by gateway while user is logged onto the database.	Once a script has been created and saved (or retrieved), it is automatically run against selected database(s). Users also have option of directly searching database without use of script.
Post-processing	Six functions enable user to translate, merge, eliminate duplicates, analyze, prepare bibliography, and conduct automatic post-processing of downloaded bibliographic citations.	Results of retrieved citations are shaped into standard format. For records obtained with BIBLIO script, transformed files can be merged, deduplicated, and sorted. Results of CHEMID script searching can be further incorporated into BIBLIO script.

6.1.2 SearchMAESTRO and Grateful MED

Both packages offer search assistance to the inexperienced end-user. SearchMAESTRO provides access to a multitude of diverse databases; Grateful MED is designed to interact with two NLM databases. Consequently, the presearch setups are different. SearchMAESTRO takes a user through simple menu selections that can be converted for use in a large number of databases; Grateful MED, which accesses one of two databases, presents an input screen that can automatically shape a user's entries into an acceptable format. Both packages can automatically log onto the selected database(s) and provide review and downloading of retrieved records. However, SearchMAESTRO offers human assistance for the end-user who runs into trouble. There are also significant differences in post-processing capabilities. At present, SearchMAESTRO does not offer additional refinement of downloaded records; Grateful MED, on the other hand, can assist with record relevancy reviews and provide suggestions for additional controlled vocabulary terms for new searching.

TABLE 6-2

GATEWAY FEATURES: SEARCHMAESTRO AND GRATEFUL MED

Component	SearchMAESTRO	Grateful MED
Presearch activities and database selection	Users are led through series of menus that progressively narrow subject of information query. Users may request access to databases they know to be pertinent to request, or have SearchMAESTRO select appropriate database, given the context of search strategy. In addition, a "scan" option is available to review search results in multiple databases.	Input search screen allows user to construct search query. Ability to use ad hoc subject terms and/or MeSH-controlled vocabulary. Users may also request direct access to NLM computer and circumvent Grateful MED input screen. User can request that search be conducted on MEDLINE, CATLINE, or both.
Search execution	Automatic log-on to selected database(s) and automatic search initiation. Ten retrieved references at a time are downloaded without intervention by user. Users can gain access to human assistant during search execution if problems are encountered.	Performs automatic log-on to selected database and begins to display retrieved results while capturing results to file on user's PC. Feedback messages on number of references located during search are provided.
Post-processing	No post-processing capabilities now offered.	Retrieved citations are placed in standard format for display and printing by user. User can make relevancy judgments about every citation retrieved. Only relevant citations need be printed out. Based on the MeSH headings attached to each relevant citation, Grateful MED can suggest additional MeSH terms for continued searching.

6.2 COMPARISON OF GATEWAY UTILITIES

6.2.1 DGIS and Micro-CSIN

DGIS, a multi-user gateway located on a minicomputer, provides a much wider array of gateway utility software than the PC-based Micro-CSIN. But the two packages handle files similarly and provide the level of support needed to create, store, and edit the files that are by-products of database searching and retrieval.

TABLE 6-3

GATEWAY UTILITIES: DGIS AND MICRO-CSIN

Component	DGIS	Micro-CSIN
Electronic mail	User can send and receive EM from other users. Private mail groups can be established.	No EM feature. However, through direct-mode searching, a user can gain access to external bulletin boards and EM systems.
File handling	Most operations – e.g., downloading data and post-processing of records – will create files. Users may modify files with choice of two editors. Files may be placed in subdirectories with hierarchical tree structure.	Every action creates a file that is stored on the user's PC. Labels every file based on activity that created file. Files can be modified and stored as either private or public files.
Subsidiary software	Users can be charged for searching activities on any registered vendor account.	As a PC software package, assumes that charges associated with database searching are for one user only.

6.2.2 SearchMAESTRO and Grateful MED

These two gateways are quite comparable in their available gateway utility features. Neither has a specific EM capability nor extensive file handling capabilities for records that are retrieved and downloaded from the databases. Neither is the subsidiary software, such as accounting, beyond the straightforward need to allow only authorized access to the databases.

TABLE 6-4

GATEWAY UTILITIES: SEARCHMAESTRO AND GRATEFUL MED

Component	SearchMAESTRO	Grateful MED
Electronic mail	None. No direct mode for dialing out and accessing other external bulletin boards or EM systems.	None. But, through direct-mode searching, a user can gain access to external bulletin boards and EM systems.
File handling	Up to 10 retrieved citations at a time can be displayed. But no files can be stored or manipulated further.	A file of downloaded NLM database records is passed without intervention by the user. This file is overwritten the next time the user completes a search.
Subsidiary software	The appropriate NTIS account is charged for database searching activities.	The user must have a password acceptable to the NLM databases. The user's NLM account is charged for searches performed.

6.3 COMPARISON OF GATEWAY SUPPORT

6.3.1 DGIS and Micro-CSIN

Because of the different environments for these two gateways, DGIS has the greater need to provide more support features because it serves a large population of users. Micro-CSIN, as a single-user gateway, does not have the same requirements to provide such data as, for example, user records. Now, however, Micro-CSIN, with context-specific messages, provides more comprehensive online help than DGIS.

TABLE 6-5

GATEWAY SUPPORT: DGIS AND MICRO-CSIN

Component	DGIS	Micro-CSIN
Help	Limited context-sensitive online help information available. User receives list of all DGIS commands with brief explanation if "help" command is used.	Three levels of online assistance provided: general, specific assistance for designated feature, and lists of available keyword and database files that can be used.
User information	One user may retrieve records of other DGIS users in order, for example, to send EM messages.	Single-user package; no directory of users.
News	News can be provided in one of two ways: as a selection from the main menu and as system news for users as they log onto DGIS.	No system or user news provided.

6.3.2 SearchMAESTRO and Grateful MED

SearchMAESTRO and Grateful MED are comparable in their gateway support features. Of greatest import is how each provides its users with context-sensitive searching assistance that is relevant and straightforward. SearchMAESTRO has the additional unique capability of providing online access to a human expert upon request.

TABLE 6-6

GATEWAY SUPPORT: SEARCHMAESTRO AND GRATEFUL MED

Component	SearchMAESTRO	Grateful MED
Help	Most menus have context-sensitive help prompt. In addition, during search sessions, users may access human expert for additional online assistance.	By pressing "HOME" key at any time during a search session, user is provided context-sensitive assistance. Searcher is then returned to the point left.
User information	No online information about other users.	Single-user package; no recording of users.
News	System news provided to users as they log on.	No system or user news provided.

6.4 COMPARISON OF GATEWAY FUNCTIONALITY

6.4.1 DGIS and Micro-CSIN

These two gateways have been developed in entirely different environments – DGIS for the multi-user population supported on minicomputers and Micro-CSIN for the single user in a the microcomputer environment. Each has its own workable approach to combining the use of both menus and commands for conducting the gateway features.

TABLE 6-7

GATEWAY FUNCTIONALITY: DGIS AND MICRO-CSIN

Component	DGIS	Micro-CSIN
Menus	Menu access provided in both up-and-down and lateral movement. Lateral movement deliberately chosen to eliminate some of the tedium of menu structure.	Menus linked hierarchically in specific sequences. Users cannot move laterally from one path to another without completing menu-driven path.
Command mode	Any DGIS menu option can also be invoked as a command. Abbreviated commands can be executed using COPS.	Menu selections can be requested as commands. The user need only enter unique set of characters, as with the COPS feature of DGIS.
Operating system and database management system	Prototype running on VAX 11/780, using UNIX operating system, INGRES database management software, and PROLOG interpreter package. DGIS software has been ported to a Pyramid 98X, an Elxsi 68000, a Gould 6050, and Sun workstations.	Current version developed under C and is designed to operate on IBM-PC/XT or AT running DOS. PC must have 512K RAM and at least a 10-megabyte hard disk. Multi-user version will be developed.
Telecommunications	Supports access from variety of terminals and PCs with standard telecommunications software and equipment. Gains access to databases with use of high-speed modems.	1200-baud modem and IBM asynchronous communications card required. Several modem types supported for dialing out to selected databases.

6.4.2 SearchMAESTRO and Grateful MED

Although these two gateways have been developed to operate in different hardware environments, each has developed a menu-driven approach for conducting database searching and retrieval. This approach is understandable since each is designed to work with the inexperienced end-user.

TABLE 6-8

GATEWAY FUNCTIONALITY: SEARCHMAESTRO AND GRATEFUL MED

Component	SearchMAESTRO	Grateful MED
Menus	User must progress through menu structures. No lateral movement available.	There are specific menu screen sequences presented to a user. Users must progress through menus before initiating or completing search.
Command mode	No commands available. All activity directed via menus.	Several elementary commands for invoking functions from main menu are available.
Operating system and database management system	Proprietary system that combines telecommunications and modem banks with storage capabilities.	Can operate on any IBM PC or PC-compatible with at least 256K RAM and one 360K double-sided disk drive. DOS version 2.0 or higher and compatible modem are required.
Telecommunications	Substantial telecommunications resources permits multiple users to dial out to wide array of databases.	Telecommunication to the NLM computer if appropriate configuration has been established.

SECTION 7

PLANNED GATEWAY ENHANCEMENTS

This section summarizes activities now under way that will enhance the four gateways – DGIS, SearchMAESTRO, Micro-CSIN, and Grateful MED.

7.1 DEFENSE GATEWAY INFORMATION SYSTEM

DTIC is now considering adoption of the commercial version of the LLNL's TIS (modified to become DGIS), which is maintained by Control Data Corporation and called ASCENT*. All the enhancements of DGIS reported in this section can be supported if a move to the commercial version of the IGP is made.

7.1.1 Directory of Resources

The DGIS online Directory of Resources will be an important asset to DGIS. There are now in development references to more than 700 DoD and commercial databases, on at least a dozen vendors, of interest to the DoD RDT&E community. The number of databases is expected to grow to more than 3,000. The directory will serve as a presearch and a database selection gateway feature for identifying, within the population of available information resources, those that are appropriate for an information query. Next will come the step of gaining access to these resources and submitting the query to each resource. The main menu presented to the user when requesting access to this directory is shown as Figure 7-1.

A prototype of the directory has been implemented by means of the INGRES relational database management system (DBMS). The directory is designed to be menu-driven, but a user can elect to search in either the menu mode or the native mode. Although there will be over two dozen searchable data elements in the record of each resource, the majority of casual or novice users are expected to search by subject, database name, or database producer. How these elements will be presented to the casual user is shown in Figure 7-2. The ability to search the directory by subject area will be a key feature, and Boolean search logic will be provided.

WHAT DO YOU WANT TO DO?

OPTION	DESCRIPTION
1	SEARCH DIRECTORY
2	DISPLAY
3	CONNECT DATABASE
4	CLEAR SEARCH SET
H	HELP
E	EXIT DIRECTORY

1 2 3 4 H E :

ENTER TOP ROW NUMBER OR FUNCTION KEY

FOR YOUR CHOICE.

FIG. 7-1. DGIS DIRECTORY OF RESOURCES MAIN MENU

WHAT WOULD YOU LIKE TO SEARCH?

OPTION	DESCRIPTION
1	DATABASE NAME
2	DATABASE PRODUCER
3	SUBJECT
H	HELP
E	EXIT DIRECTORY
B	BACK UP 1 SCREEN

1 2 3 H E B :

ENTER TOP ROW NUMBER OR FUNCTION KEY

FOR YOUR CHOICE.

FIG. 7-2. DGIS DIRECTORY OF RESOURCES SEARCH MENU

DTIC plans to incorporate a natural language interface in the future. Eventually, the directory and CCL – see below – are expected to interface closely.

7.1.2 Common Command Language

The CCL project began with the realization that DGIS could take a user to a database, but that the user was then left to his/her own devices. This approach is acceptable if only a limited number of databases are to be accessed and the user is willing to learn the command structures for every one. However, the potential of gaining access to dozens of pertinent databases via DGIS confronts the user with the complex task of learning the searching features of several database system vendors – e.g., DIALOG, DROLS, ORBIT, NASA, and BRS. This presearch strategy capability, which is used during search execution, will become increasingly necessary as more end-users are provided access to DGIS.

DTIC, therefore, targeted several information systems that are probably of general interest to the DoD information community and structured a DGIS CCL activity to develop a CCL in an incremental fashion. The newest CCL project objective is to "design and implement a DGIS CCL System (CCLS) that incorporates AI tools for proficiently searching and retrieving in heterogeneous information systems in a standard manner, for use by the professional searcher and the knowledgeable end-user."

Basically, the CCL is a standardized command language replacement for diverse native command languages. This means that the user learns one command language, which can be processed by CCL, to search diverse databases. DGIS CCL follows the draft standard developed by the National Information Standards Organization (NISO) for the CCL command set.

The first phase of the CCLS has resulted in prototype CCLs for DIALOG, BRS, NASA/RECON, and DROLS TR. Those four prototypes were programmed in C. The process gave DTIC valuable experience and insights into the CCL concept. The prototype review not only defined problem areas but provided the basis for DTIC's decision to apply AI tools for the next phase of the CCLS, which will expand in providing a user-system interaction for accessing information.

There are five modules for planned CCLS development:

1. Artificial Intelligence Management Module. This module identified applications and requirements for the AI-based CCLS. Because CCLS will interface users with heterogeneous information systems, this module will be responsible for the acquisition of an AI system, comprising both hardware and software, for continued CCLS development. CCLS, already in progress with PROLOG programming, will expand to incorporate knowledge bases and become an expert system for assisting with database searches. An AI system will be needed as a means of increasing development productivity and proficiency in expert system development. For the CCLS, the AI system will establish intelligent programs that make the human-machine interaction more human-like. In addition, the AI system will be networked to DGIS to serve all of DTIC's AI development projects.

2. Requirements Management Module. This module combines three basic activities: (a) researching and mapping database native command languages to the DGIS CCLS, (b) identifying the individual operating characteristics of information systems for integration into the DGIS CCLS, and (c) building the Command Language Knowledge Base for the DGIS CCLS.

   The databases targeted for the module are:

   DROLS [TR, CF, Work Unit Information System (WUIS)]
   Manpower and Training Research Information System (MATRIS)
   NASA/RECON
   DIALOG (which now includes DoE/Energy as the DoE STI source)
   BRS
   ORBIT.

   Successful development of CCLS for these information systems will allow standardized access to about 600 DoD and commercial databases, as identified by the Directory of Resources.

3. Artificial Intelligence Engineering Module. This module is concerned with the technical programming and AI engineering of CCLS. The CCLS will be a melding of PROLOG and C programming. PROLOG will drive the CCLS functions, and C will support the features on the DGIS that will enable the user to make use of AI-based CCLS. This activity will also be concerned with integration of AI hardware and software tools into DGIS.

4. CCL Standards Module. This activity will track national and international CCL standards activities for the CCLS effort. The CCLS development will follow closely the activity of the NISO and its Z-39 Group for Common Command Language Standardization.

5. Production CCLS Engineering Module. As the CCLS becomes stabilized, provision for a production version will be formulated to insure continuity.

A final report of the project will lay the basis and state the requirements for continued CCLS engineering.

With the transition to an AI-based CCLS, the goals of the project will incorporate AI-driven capabilities as follows:

- Standardize a command language to communicate with bibliographic databases.
- Create a CCLS that helps the user in searching unfamiliar database systems.
- Provide the means for easy-to-use, intelligent, and productive search sessions.
- Remain flexible enough to adapt to changes and enhancements.

7.1.3 Post-Processing Enhancements

A major effort was undertaken to modularize and reduce the code involved in the existing DGIS post-processing routines, while adding new capabilities. DTIC expects that in the near future, several new enhancements will be brought into the DGIS post-processing prototype. These are the main highlights of these enhancements:

- In checking text for duplication, post-processing will deal equally with lower-case and capital letters. At present, duplicates are not recognized unless all letters match exactly.
- In checking text for duplication, post-processing will match an author's first name or initials in one entry against that author's initials or first name in other entries. At present, duplicates are not recognized unless the entries are identical.
- If the downloaded bibliographic records include parenthetical information, such text can be ignored during duplicate identification, but retained for review.
- Once a group of downloaded records is reviewed for duplicates, the longer version of the record – presumably, the one that contains more information – will be retained instead of any shorter duplicate records.

More context-sensitive messages will be made available so that the user is given more specific guidance on what problem has arisen and how it can be corrected.

DTIC has also begun considering how to perform post-processing on full-text records and numeric data. DGIS now deletes all but the first six lines and the last six of a full-text record that is post-processed.

7.1.4 User Chargeback Services

The User Chargeback Services will be offered to DGIS users by the end of calendar year 1988. The object of the service is to establish a single, streamlined system of DTIC contractual arrangements and payment procedures that will ease access by DGIS users to fee-based online domestic and foreign commercial databases and document delivery services. The service will establish accounts on behalf of users for commercial databases, commercial document-delivery services, and online searching activities performed with SearchMAESTRO. Charges that a user incurs while using any one of these services will be charged back to the NTIS account of the user's organization.

Test sites for SearchMAESTRO chargeback services were in place in November 1987. These services will be made generally available to registered DGIS users who have filed special forms and have provided NTIS account information.

7.1.5 Access to More Databases

DTIC will continue to identify additional databases and services, particularly in the fields of military science and foreign technology that are deemed appropriate to its RDT&E user group. DTIC plans to make them available through DGIS when identification has been made.

7.2 SEARCHMAESTRO

The two enhancements described next reflect only activities that are slated for the version of the EasyNet gateway, known as SearchMAESTRO, that has been adapted for DoD users. These two enhancements do not encompass all features under development by Telebase for EasyNet. or other specialized versions of the gateway for other clients.

7.2.1 Save Our Search Service

Telebase now offers a free service that enables users to receive online human assistance while they are connected to the EasyNet system. This service, known as

SOS, will be made available to SearchMAESTRO users. Telebase itself staffs the SOS service. It is possible, however, to direct SOS inquiries to sites other than Telebase. This can be done on either a selective or a system-wide basis. DTIC has developed functional requirements for a SearchMAESTRO system enhancement through which selected SOS calls may be channeled to sites determined by DTIC. This enhancement will be made available to users of SearchMAESTRO only.

The proposed system will route SOS inquiries selectively to the EasyNet staff or to a site selected by DTIC. Selection criteria will be based on the user's response to a prompt at the time the request for "SOS" is made. Initially, SOS calls for DTIC assistance will be routed to DTIC at the Cameron Station offices. Two SOS stations will be provided, enabling two staff members to respond simultaneously to users' queries concerning DROLS.

In the future, DTIC may set up additional SOS sites. At that time, DTIC may want site selection to be determined by the subject of the request. The user may respond to a subject menu so that the Telebase facility can determine to which DTIC site – for example, an Information Analysis Center (IAC) – the SOS call should be directed.

7.2.2 Access to Government Research Databases

During initial development of SearchMAESTRO, DTIC staff showed interest in creation of a menu option that would offer SearchMAESTRO users access to these additional Federal Government research and technical databases: DTIC TR, NTIS, DoE Energy, Aerospace Database, DTIC WUIS, and Federal Research in Progress (FEDRIP).

EasyNet has demonstrated the "scanning" capability that enables users to search for and retrieve records with a common search term, without any need to know the different command languages of the various databases (see Section 3.1.1). DTIC wishes to apply this "scanning" capability to the Federal Government databases specified above. SearchMAESTRO will perform the search, report the number of records found within each database, and provide the user with the option to request the retrieval of records from any one or more of the databases searched. The "scan" feature will be included as a menu option under the Government Research Database menu selection.

7.3 MICRO-CSIN

7.3.1 Incorporation of Grateful MED Features and MeSH Lookup Functions

NLM is testing a version of Micro-CSIN that can invoke the Grateful MED software package. Now a Micro-CSIN user who wishes to search the NLM MEDLINE or CATLINE databases will be automatically transferred to the Grateful MED main menu. Moreover, Grateful MED offers a useful feature that enables a user to review and either save or discard retrieved citations and then proposes additional MeSH terms that may be appropriate for continued searching (see Section 5.1.3). The same capability will become available to users of the BIBLIO script in Micro-CSIN. Users will be able to save the list of proposed MeSH terms and use it in a BIBLIO script.

7.3.2 Multi-User CSIN

An earlier version of a multi-user CSIN package was developed (see Figure 4-15) on a VAX 11/780 and offered the CHEMID script only. Since that initial multi-user development, the next few versions of CSIN were single-user packages. NLM is now beginning to explore the most efficient environment – probably the 80386 chip technology with the UNIX operating system – for reintroducing a multi-user version of CSIN and creating a multi-user package again.

7.3.3 Direct Connect to the American Medical Association Network

The American Medical Association (AMA) through its subsidiary, American Medical Computing, has provided the medical community with a variety of services that include clinically and medically oriented databases, EM, and electronic bulletin boards. The AMA has selected SoftSearch, Inc., to provide the telecommunications network support for these services. The SoftSearch facilities include both the Digital Equipment Corporation (DEC) TOPS-20 KL-2060 systems and DEC VAX 8500 systems operating under VMS 4.X. In addition to its own network, SoftSearch uses American Telephone and Telegraph (AT&T) WATS, CompuServe (via X.25), and Telenet (also via X.25) to make outgoing calls or accept incoming ones. To further expand database access for Micro-CSIN users, a direct-access connection to the AMA Network (AMANET) will be created by means of the Micro-CSIN "Direct Connection" option. Ultimately, ability to access AMANET and use the full script

interaction and post-processing features of Micro-CSIN while interacting with AMANET resources will be provided.

7.3.4 Agency for Toxic Substances and Disease Registry Workstation

The Micro-CSIN workstation enables users to search for chemical identifying information if a registry number, synonym, or proper name is known. Plans are now underway to augment this identification process with the capability to base a search on physical or related properties. The Agency for Toxic Substances and Disease Registry (ATSDR) emergency response personnel will then be able to look up a chemical entirely on the basis of physical properties; these can often be ascertained first. The purpose of this new script for the emergency personnel will be to identify the precise chemical(s) involved in a spill via gross observations.

7.3.5 Compact Disc Read-Only Memory Databases Linked with Workstation

As more and more medical and biotechnology information resources are transferred to the compact disc read-only memory (CD-ROM) format, NLM will explore ways of providing access to them from the Micro-CSIN workstation.

7.4 GRATEFUL MED

One overall modification of Grateful MED will be production of a version in the C language. This modification will make the software package more portable. Two additional enhancements now under consideration are to enable Grateful MED to extend its searching capabilities into chemical databases and to permit the scanning of the MEDLINE backfiles while retaining a search that was performed on the current MEDLINE database.

7.5 NLM APPLICATION GATEWAY WORK STATEMENT

In 1985, NLM sponsored panel discussions on methods of improving and extending the use, delivery, and computer-based manipulation of biotechnology information for the research and medical practice communities. In 1986, a new concept called Advanced, Biomedical Information Databand Exchange (ABIDE) was developed in support of the Congressionally mandated "Biotechnology gain Initiative." To lend further impetus toward developing a technical approach to gain comprehensive access to biotechnology information, legislative action in the form of

H.R. 5271 established a National Center for Biotechnology Information at NLM at the end of 1986.

Among the elements of the ABIDE framework, the concept of "gateway" plays a prominent role. NLM expects the gateway to provide a point of access, a set of access methods, semantic and syntactic conversion functions, and possibly links from users to all NLM or other biomedical information services of interest. Among the goals of the ABIDE program, two are of particular pertinence in this context:

- Provide concurrent, switchable access to multiple information systems

  The first goal is motivated by users' frequent need to access several information services to carry out an effective search, without losing the context of the search on any one database. Moreover, some situations require that information obtained from one search be incorporated into the query presented to the next, as the search target is narrowed.

- Support terminal and PC-based access to information systems

  The second goal is motivated by the large base of users of biomedical information still using terminals and, for example, public data nets or direct dial to reach specific services.

NLM has concluded that it is essential to provide a base for supporting both terminal and PC-based use of information services to create a seamless path of migration from terminal to workstation-oriented service. This approach is to counter a frequent situation in which users find it essential to access both the MEDLINE facilities and the TOXNET system. Technology now available requires that such users engage in separate search sessions with these databases, though additional knowledge could, in fact, be gained by switching back and forth among database services, saving specific searches, and processing the results of each session.

In addition, NLM needs to communicate with other national libraries and service organizations to exchange or share bibliographic information. NLM plans to accommodate the "Linked Systems Project: Standard Network Interconnection" that has been developed by the Library of Congress, Western Library Network, and Research Libraries Group, Inc.

Finally, the National Institutes of Health (NIH) have acquired and installed very-high-power computing resources, such as the CRAY, DEC 11/785, and

8600 facility at the Advanced Super Computer Laboratory (ASCL) in Frederick, Md.; this should be made accessible to a geographically dispersed community of researchers, both within NIH and outside.

This gateway base, as now planned, will be created as a multi-user and concurrent multi-session system on a UNIX-based (UNIX 4.3 BSD) application gateway (AG), potentially supported by DEC MicroVAX II hardware. NLM is seriously considering – as is DTIC – use of the commercial version of LLNL's TIS from CDC, known as ASCENT* (see Section 7.1), to serve as the gateway's foundation.

The AG will be the primary interface to all or selected NLM information services. It will be the target of users' calls via Tymnet, Telenet, and the local area networks (LANs) at NLM and Advanced Research Projects Agency Network (ARPANET). In turn, the AG, after interacting with the caller, will call the desired services. What will make the AG of special interest will be its ability to maintain several concurrent outgoing connections for a given incoming user. As an example, a version of Micro-CSIN could be created to operate in the AG. The software could be invoked if the user indicated the need to access the NLM facility. Such implementation would bring many of the benefits of the microcomputer version of CSIN to a user who might have only a simple terminal. In a more elaborate scenario, a user with a PC might be running the Micro-CSIN software, suitably configured to interact with the AG. Such an interaction would permit the AG to assist the user with the management of multiple links to multiple services but not be required to support Micro-CSIN functionality. NLM is considering ways of making available gateway functionality, whether on a centralized basis, by distributing PC or intelligent workstation software, or both.

NLM personnel are also working on approximately 20 interim projects, grouped into nine broad categories, that will enhance the ABIDE gateway vision to encompass the following: online expert help (as with the SOS SearchMAESTRO capabilities), meta-browsing (similar to the DGIS Directory of Resources), query language translation (similar to the DGIS CCLS project), deferred services for low-priority access, local data manipulation for citation processing, support for personal bibliographic and native databases, presentation enhancement software with available graphics and color features, multiple access to a second computer based on

knowledge acquired during sessions with an initially-accessed computer, and feedback data for NLM management needs.

In summary, ABIDE envisions an architecture where, in response to a user's need for information, powerful PCs at the workstation will create programs for distribution among the multiprocessor gateway computers that provide for access to existing databanks.

SECTION 8

REVIEW OF THE CAPABILITIES OF OTHER GATEWAYS

The gateways reviewed in this section are offered as examples of approaches that have been developed in the two hardware environments – mainframe/ minicomputer and PC. They do not cover the spectrum of gateways available. Rather, they were chosen for review because of their potential applicability to DTIC and NLM. Some of these gateways may never become actual products but some of the design issues and approaches have been reviewed for use in enhancing DTIC and NLM gateways.

8.1 MAINFRAME ENVIRONMENT

The four mainframe gateway projects described in this section are primarily experimental prototypes. They seek to probe further into the following questions: How does an individual identify an information need? How does a user formulate a search strategy to obtain references to relevant information? How can a user refine the search strategy until it is acceptable? How can a user review the results of the retrieved information? For the most part, these gateways concentrate on presearch strategy and database selection.

8.1.1 Connector for Networked Information Transfer

The Massachusetts Institute of Technology's (MIT's) Connector for Networked Information Transfer (CONIT) gateway project was implemented on the MIT Multics computer system. The gateway was developed at MIT's Laboratory for Information and Decision Systems with research grants from the National Science Foundation, NLM, and DTIC. The system provides access to three major retrieval systems – DIALOG, ORBIT, and NLM. This gateway is the root of a "family tree" whose progeny can be traced to the Individualized Instruction for Data Access (IIDA) project (see Section 8.1.2) and its derivative, Online Access to Knowledge (OAK), (which is discussed in Section 8.3.3). The three systems are similar in design, but each has contributed uniquely to the thinking in gateway design.

The CONIT gateway approach employs AI techniques and is based on the concept of a system in which communication among users, retrieval systems, and the gateway are defined by a set of rules. That approach makes for a compact – yet flexible and modular – description of the message-handling protocols required in the human-computer interface. The approach helps interpret the messages and the interactions that take place between the user and the system. This message interpreter, which is the heart of the gateway, provides a high-level, application-oriented language in which CONIT can tell the user, in a heterogeneous and interactive manner, how to interface.

The main emphasis in the development of CONIT has been on objective evaluations of how effective intermediary assistance (or gateway) systems are and how effective they can be. This experimental approach has allowed for further development and refinement as new information is acquired on the gateway's performance.

The simplified command/argument language, together with several natural language features, are key to the CONIT system. CONIT takes each word in the query phrase entered by the user and derives a root form for it. This approach was chosen because it avoided the major problems posed by the other dominant gateway designs, specifically, the standardization of searching across the accessed systems. Instead, the CONIT design has sought to identify the desirable features of a common information retrieval language and associated computer-assisted instruction for the user interface. The concept was to build an interface that provided the following four elements: use of common English words as terms for commands and search strategy arguments; use of English-like constructions for the commands and arguments; provision of English responses to the users; and use of a natural-language approach to a common indexing and search vocabulary.

In helping a user to plan a search, CONIT begins by asking eight questions about the information need or problem. First, the user is asked to assign a title to the search. Next, the user is asked to describe the search problem as completely as possible. To set ceilings when the user is logged onto a selected database(s), maximum time and dollar limits are then established for the search session.

The next questions are concerned with the following areas: (1) how many documents the user believes have been published that address the information

question, (2) how many relevant documents the user is aware of, and (3) how many documents the user wishes to identify as a result of the search. Next, the user indicates the types of documents that will be most useful – professional journal articles, conference papers, books and monographs, reports, theses, magazines and news articles, or all of them. CONIT then confirms the responses to these eight questions and proceeds to the next stage of search planning.

This stage consists of specifying the most appropriate databases for the search. CONIT can help by listing files in subject areas or subareas. In a series of menus – similar to the database selection menus available in Micro-CSIN – the user chooses the databases that appear to be most relevant and designates them by the database codes used by CONIT.

The next stage requires the user to prepare a search statement. Three types of search statements are offered by CONIT: topic, author, and citation. This approach is also mirrored in the OAK gateway. To create a search strategy, the user is asked to formulate a "keyword phrase," which is a phrase of one or more words, in which each word represents an important concept in the search topic. Here the user can opt to use the search title established in the first stage or introduce new words to create a keyword phrase. From these keywords, CONIT then constructs a search strategy that can be applied across all the selected databases. For example, if the user enters "intelligence retrieval assistance," CONIT reconstructs it into the search strategy illustrated in Figure 8-1.

r4 (s4) = r1 AND r2 AND r3
where r3 (s3) = assistance
r2 (s2) = retrieval
r1 (s1) = intelligence

FIG. 8-1. CONIT SEARCH STRATEGY

CONIT then suggests that the user submit the search to the selected databases and receive some feedback before modifying the search strategy. Several review and evaluation options are now available. For example, CONIT can provide a user with a review of the comprehensiveness of the search that has been created, in addition to modifying the problem description, qualifiers, and strategies.

If the user chooses to submit the search strategy to the first database selected, CONIT performs all of the logging-on and search execution steps without any need for the user to intervene. The results of the search are displayed as they are retrieved, as shown in Figure 8-2.

```
You are now connected to the Lockheed DIALOG retrieval system -
You are now connected to the REPORTS (Code Name UR) file.

Search s1 (find intellig:) found 9989 documents.
Search s2 (find retriev:) found 11693 documents.
Search s3 (find assist:) found 27906 documents.
Search s4 (combine s1 AND s2 AND s3) found 23 documents

Your search strategy has been run.

Your options now are:

 + + 1)  See information about retrieved documents.
 + + 2)  Run search in a different file.
 + + 3)  Narrow search: find fewer documents.
 + + 4)  Modify search in some other way.
```

FIG. 8-2. CONIT SEARCH EXECUTION AND RETRIEVAL RESULTS

If the user wishes to see information about retrieved documents, three options are available – standard citation format, titles only, or all available information. Both online display and offline printouts are presented. If the user wishes to sample the group of retrieved documents, to determine whether the set is truly pertinent, suitable commands are available.

CONIT provides an extensive group of commands for modifying the search strategy. Modifications to broaden a search, narrow a search, or rerun the search entirely are offered as options. Complete explanations of these commands and their

ramifications are available. For example, if the user chooses to broaden the search strategy, the following options are displayed:

HELP BROADEN

To broaden your search and find additional (hopefully relevant) documents, there are several techniques you can try.

Your options now are:

+ + 1) Add alternate (synonymous) term(s) for a search.
+ + 2) Loosen a search by dropping a factor (use fewer ANDs).
+ + 3) Relax combination requirements (change ANDs to ORs) - this means retrieving documents from EITHER one OR another search instead of requiring the document to be in both searches.
+ + 4) Replace search term with better term(s).
+ + 5) Replace exact, untruncated, or long terms with words or shorter truncated stem searches.
+ + 6) Replace phrase of proximity searches with broader proximity searches.
+ + 7) You may need to identify better files to search in.
+ + 8) Do an entirely new search or combination of searches.

You can explore and make several changes before rerunning your searches. Type an option number or any CONIT command.

FIG. 8-3. CONIT SEARCH STRATEGY MODIFICATION

Users can also check the effectiveness of their search strategies by evaluating the recall, precision, and cost associated with executing the search. The options offered if the user wishes to evaluate the search strategy are as presented in Figure 8-4.

Your options now are:

+ + 1) Estimate recall of a search.
+ + 2) See why a search cannot retrieve every relevant document.
+ + 3) See information about three different recall estimation mechanisms.
+ + 4) Do something else.

Type an option number or any CONIT command.

FIG. 8-4. OPTIONS TO GAUGE THE EFFECTIVENESS OF CONIT SEARCH STRATEGY

Over the years, several experiments have been conducted to assess the effectiveness of CONIT for end-users. Reporting on work accomplished in 1983, Richard Marcus, the "father" of the CONIT gateway, reported:

> Experiments with one version of CONIT (identified as CONIT 3) have shown that it is possible for computer intermediary systems to assist end users, who had no previous experience in operating retrieval systems, to obtain in this way information they needed from dozens of heterogeneous databases on four different systems. In these experiments users found relevant information typically beginning within 20 – 30 minutes of their online session; instruction was provided entirely by the intermediary system – no additional human assistance was necessary (except to handle certain system problems). . . .
>
> Along with the positive achievements demonstrated through CONIT 3, several questions were raised. For example, although all experimental users were able to retrieve *some* relevant documents, the recall levels were determined to be rather low: from 0.2 down to 0.01. Furthermore, while session times – and other performance measures – were acceptable to – indeed, praised by – the experimental users, the question was raised as to whether human expert information specialists acting as intermediaries could achieve greater effectiveness in shorter time. Finally, there was the question of whether a computer intermediary system could be made sufficiently effective so that its performance would equal or be superior to that of human experts in all respects [1]

This last question is one that has dominated the CONIT gateway project, as well as all other gateway design and development approaches.

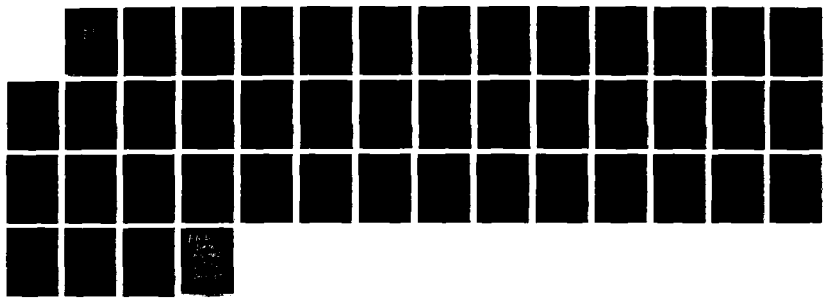
The CONIT experiment can provide valuable guidance to the gateway plans at both DTIC and NLM. CONIT has developed a utilitarian bridge between diverse databases and the user by establishing a dialogue that can be understood by both parties. For the user this dialogue appears to consist of straightforward English commands and statements. Users have available a plethora of options that permit search strategy modification, display and printing of search results, and review of search strategy effectiveness. The user is not only presented with sorted and merged post-processed results, but instructed in how to conduct a more effective search. The combination of presearch strategy formulation with post-search strategy reformulation is a powerful tool for the user.

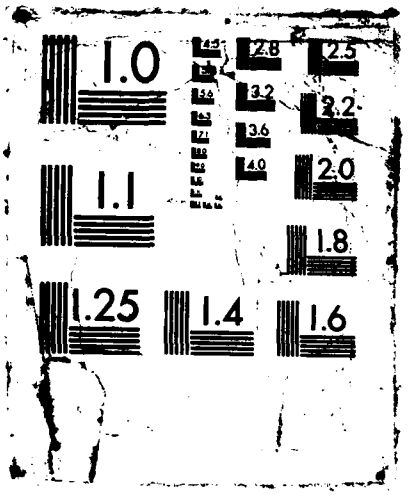
---

[1] Marcus, Richard S. "An Experimental Comparison of the Effectiveness of Computers and Humans as Search Intermediaries." *Journal of the American Society for Information Science* (1983): pp. 381 – 382.

AD-A193 362 CAPITALIZING ON EXPERIENCE WITH INTELLIGENCE GATEWAY SOFTWARE(U) LOGISTICS MANAGEMENT INST BETHESDA MD C W SHOCKLEY JAN 88 LMI-DL604R1 DTIC-TR-88/7 2/2

UNCLASSIFIED MDA903-85-C-0139 F/G 12/5 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

8.1.2 Individualized Instruction for Data Access

Beginning in 1976, a project involving Charles Meadow, who later participated in the OAK project (see Section 8.3.3), was undertaken by the Drexel University School of Library and Information Science and the Franklin Research Center, with the objective of developing a computer intermediary to aid scientific and technical end-users of online bibliographic information retrieval systems in performing database searches.

The IIDA project was designed with a specific user population in mind. Although other groups would find it useful, the design was predicated on the following user characteristics: (1) IIDA users are end-users of the technical literature and are aware of a specialized professional vocabulary; (2) they are aware of the professional literature and the concept of citations and abstracts; (3) they are not computer-shy or anxious about hands-on use; (4) they need to locate information and have serious information queries; (5) they are expected to use IIDA 1 to 10 times a year; (6) generally, they want a good citation, not an exhaustive search; and (7) they are typically nonusers of libraries – or reluctant users at best.

IIDA software made extensive use of software developed for the CONIT project. The IIDA computer was actually the MIT Multics computer. Users dialed IIDA, via Telenet, and logged onto the Multics computer at MIT. Once logged on, users received a computer-assisted instruction program. Its purpose was to introduce the student user to the minimal set of DIALOG commands necessary for a meaningful search. Only when they used IIDA did users deal with the DIALOG language and commands. After the learning program was completed, users were encouraged to try DIALOG searches on their own. A search topic was suggested when the user was logged onto DIALOG; through an interactive presentation, more complex search commands were presented and reviewed.

To detect errors and poor uses of the DIALOG system, IIDA employed a set of diagnostic procedures, grouped into three classes:

- Syntactic

  IIDA would detect any error in DIALOG syntax and provide the user with information about the nature of the error. These error messages were clear statements of what was acceptable in the command and what was not.

- Procedural error – single command

  These diagnostics apply to commands that are syntactically acceptable but, in light of the search strategy's history, are found to be faulty in some way. These faults can be of the fatal type (e.g., "COMBINE 3 AND 4," where set 4 has not been defined), or minor faults, such as repetition of a command.

- Procedural error – strings

  These strings were defined as an unbroken series of commands of the same type. For instance, if a user enters more than five consecutive COMBINE commands, a diagnostic is prompted, and IIDA suggests to the user that a different command should be used.

IIDA, like CONIT, was designed to make it possible for an end-user to perform online bibliographic information searching without the help of a trained intermediary. CONIT provides its own CCL, which the end-user employs to search several databases supplied by more than one commercial vendor and, therefore, requires the ability to parse and interpret more than one access command language. IIDA, on the other hand, teaches a single language and offers the user a limited choice of files for searching.

As a learning tool, the IIDA project experience can provide both DTIC and NLM with practical guidance, when each contemplates the level of knowledge about database searching and retrieval that should be imparted to a novice user who needs to search STI. Such instructional capabilities will be required as gateways enable more and more end-users to gain access to databases without human intermediaries.

8.1.3 Composite Document Expert/Extended/Effective Retrieval

The Composite Document Expert/extended/effective Retrieval (CODER) project at Virginia Tech was designed as a prototype for an intelligent information retrieval system, as well as an experimental test bed. Acting as a front-end to the menu-driven Virginia Tech Library System, CODER, which was developed under UNIX in PROLOG and C, was to answer such fundamental questions as these:

- Can an intelligent automatic intermediary serve users as well as (or better than) a human?

- Can a knowledge of user characteristics provide the foundation for the development of an interactive human-computer dialog that contributes to more effective retrieval?

- **In the environment of full-text databases, will an intelligent gateway be able to achieve a higher level of recall and reasonable precision?**
- **Have natural language processing and knowledge representation methods matured enough for better representations of documents to be presented to the user for heterogeneous databases?**
- **Is the PROLOG language, that is now used by DTIC for its CCL project, appropriate for developing an experimental information intermediary?**
- **Can several machine-readable dictionaries be used and analyzed to aid in document and query processing?**

The design of CODER, which was derived from Gerald Salton's experimental SMART system at Cornell, is also similar to that of Intelligent Intermediary for Information Retrieval (I^{3}R), but there are some differences. CODER sought to address the overall retrieval strategy problem – not just improve retrieval – and continue to use natural language processing. The overall structure of CODER is shown in Figure 8-5.

Key to the goal of the CODER experiment is investigating the value of natural language processing of documents and queries with the CODER lexicon. One part of the lexicon is obtained from machine-readable text that is fed into the gateway (or domain knowledge) while the other portion is domain-independent. The CODER system stores the lexicon in the form of PROLOG facts so that machine-readable material can be interpreted for the user during the retrieval process. The object is to continue to explore the relationship of words in documents, as well as ways of indexing or classifying them to meet the searching needs of end-users.

If either DTIC or NLM considers the possibility of employing two controlled vocabularies – one for indexers and one for end-users (see Section 9.1.3) – the experience gained by the CODER project in building vocabularies from the full text of the materials that compose the database may help.

8.1.4 Intelligent Intermediary for Information Retrieval

The I^{3}R project at the University of Massachusetts is a document retrieval system that combines use of domain knowledge, exhibited by the expert intermediary interface, with statistical techniques. A typical session using I^{3}R involves three major phases: query formulation and refinement, search execution,

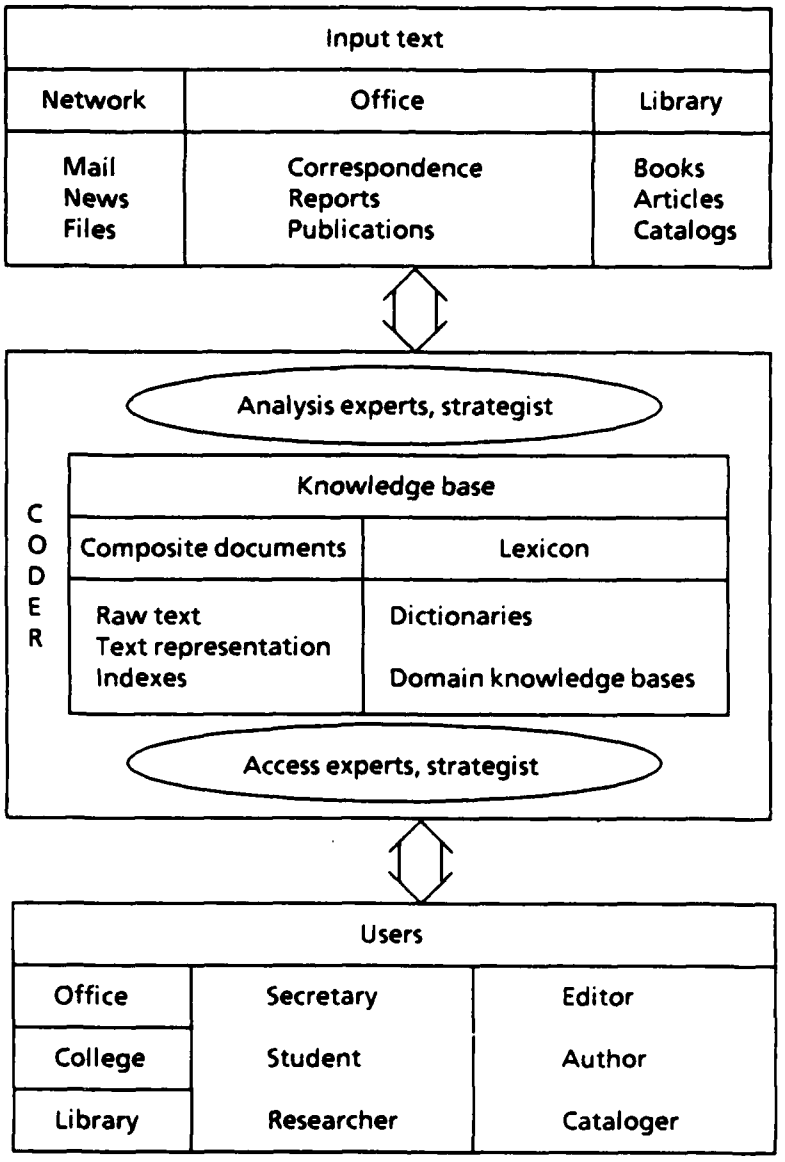


FIG. 8-5. OVERALL STRUCTURE OF CODER

and user evaluation. The purpose of the design was to produce a session of retrieval activities and user evaluation that would be similar to a human intermediary.

The I^{3}R system is composed of six components: user model, request model, domain knowledge, search controller, browsing expert, and the "explainer" or help messages. During query formulation and refinement, I^{3}R attempts to build an accurate description of the information need. This description is referred to as the "request" model. The request model contains a list of concepts and their relative

importance. These concepts are derived from the user's query and domain knowledge. Domain knowledge can be acquired during the development of the request. The knowledge that is specific to an individual user is retained as part of the "user" model.

The information in the request model is used by the retrieval strategies to retrieve ranked lists of documents. Since retrieval is viewed as an inference process, the ranking is determined by the statistical certainty of the inference for each document contained in the database.

In the evaluation phase, the user reviews the lists of pertinent documents, identifying both documents and parts of documents such as phrases, that are important. This evaluation feedback is used to update the request model. Evaluation is conducted in conjunction with a browsing capability, which enables the user to look over documents and other information in the database.

The system architecture that supports I^{3}R is similar to the CODER project in that the gateway builds a knowledge base and provides the interface between material that is to be made accessible and the users who wish to retrieve information. A prototype of the I^{3}R system is being implemented in the CommonLisp and C on a Micro Vax-II workstation. This workstation provides the high-resolution graphics capability and processing power required for the system.

As with CONIT, rules are established, reviewing the messages between the system and the user to determine what action has occurred or should occur. One major contribution of the I^{3}R project to gateway enhancement consideration by DTIC and NLM is the browsing feedback, which modifies the "request model" or search strategy. While browsing, the user identifies documents, terms, or concepts to refine the request model. A browsing window displays a document "node" and shows the user how to navigate through the knowledge base to continue examining other linked documents. This powerful retrieval concept should be considered by DTIC and NLM, because I^{3}R can interpret the user's need from data acquired from the user model and point the user to a network of related documents that may not have been identified specifically in the request model.

8.2 LOCAL WORKSTATION OR PC ENVIRONMENT – ACCESSING HETEROGENEOUS DATABASES

As with SearchMAESTRO and Micro-CSIN, these gateways typically provide search guidance and access to heterogeneous databases supported by different vendors.

8.2.1 DIALOG Business Connection and DIALOGLINK

The DIALOG Business Connection is a menu-based package designed for business professionals who have recurring needs for information about companies, products, and markets. It combines DIALOG's search techniques with an intelligent front-end to the DIALOG databases. DIALOG Business Connection is composed of five main sections, called "applications," which correspond to five major needs of the business professional as determined by DIALOG:

- Corporate intelligence
- Financial screening
- Products and markets
- Sales prospecting
- Travel planning.

For each application there is a customized set of menus and prompts that lead the business professional to define the information needed. Based on the responses, the DIALOG Business Connection software identifies the database sources that may contain relevant citations.

The software of the DIALOG Business Connection formulates and enters the search commands for the user and then offers a choice of display options, such as a "Share of Market Report" for an industry, mailing labels for identified sales prospects, or summaries of articles from the trade literature. DIALOG Business Connection includes access to the following databases:

- Business Research Corporation
- Disclosure Information Group
- Dun and Bradstreet, Inc.

- Market Data Retrieval, Inc.
- Media General Financial Services, Inc.
- Moody's Investors Service, Inc.
- Official Airline Guides, Inc.
- Predicasts
- Standard & Poor's Corporation
- Thomas Publishing Company, Inc.
- Trinet, Inc.

The DIALOG Business Connection provides menus and prompts for the new or occasional end-user and commands for more experienced searchers.

DIALOGLINK is DIALOG's communication and accounting software. It provides such features as creation and editing of search strategies before log-on to the database(s). DIALOGLINK also has the ability to save, redisplay, and print information already received; local help functions; the ability to recall and modify the last command sent to the database; and accounting features that enable the user to create lists of searching costs by searcher, client name, client code, etc.

DIALOG recently announced an online enhancement for users logged onto the system called OneSearch. OneSearch enables online users to search as many as 20 DIALOG databases with a single command that combines results and output. OneSearch enables searchers to compare results across databases, as well as to modify and combine search terms and to display results.

The contribution of these DIALOG packages to the work contemplated by DTIC and NLM is the shifting of command complexities and menu presentations to reflect the user type. The accounting features may be useful for DTIC as the user chargeback service becomes available. Finally, for both DTIC and NLM, the capability of comparing databases retrieval results, as provided by OneSearch, can be of benefit to users who may wish to refine their database access lists for future searching needs.

8.2.2 Sci-Mate™

In 1978, the Institute for Scientific Information (ISI) decided to create a way for librarians and scientists to solve one of their perennial problems – managing the flood of reprints that they use in their daily work. The original idea was for scientists or librarians to use a microcomputer to help them keep track of books, papers, pamphlets, correspondence, and other kinds of text.

The initial product, Sci-Mate, had two major components. First, it contained a universal online searcher, a menu-driven system that permitted searches to be submitted in plain English to major commercial databases. Each step of the process was fully prompted and easy to follow. For example, as the user changed database selections, a new set of field names corresponding to the database record was displayed. On-screen prompts helped the user work with the display to construct an efficient database search. Second, it provided an offline personal data manager. This feature enabled a user to create a local database of citations that could then be searched. The two components were designed to work together. Citations retrieved during database searches could be offloaded into the personal database and could then be stored, sorted, edited, and later, reviewed. Downloaded records could be merged with locally created records.

ISI Publishing, a division of ISI, announced the release of Sci-Mate version 2.1 in the summer of 1987. This is a performance upgrade of the original Sci-Mate Software System. With this version is Editor, a tool for formatting references automatically. The Manager can accommodate files up to 8 megabytes and offers searching and sorting of the downloaded files.

Sci-Mate makes possible the choice of searching BRS, DIALOG, NLM, and SDC via a common menu-driven interface or the option of searching these systems in their own native languages. The user may also search in a passive terminal mode to access other host systems. On Sci-Mate, at each step in the search, the user can return to an earlier step or have online help screens displayed by entering a question mark. As a search aid, Sci-Mate has a browse function with which the user can expand – into the online index, for example – to find variant forms of an author's name. For the correct use of field tags, the user needs to consult the documentation of the online system and specific database.

Because Sci-Mate deals with several online systems, saved strategies can only be run on compatible files on the same online system. When saving a strategy, Sci-Mate gives the user the option of retaining the entire strategy or just the portion that proved most useful.

Downloading is probably the strongest feature of the Sci-Mate software when the Personal Data Manager is available to the user. On Sci-Mate, retrieved records are first input into a temporary workfile from which they can be transferred to the Personal Data Manager. With the Display/Offload option, the user can transfer all, none, or selected records. With the Personal Data Manager, citations are structured immediately according to templates, which the user has the option to redesign. The user can also add records, search the records in the personal database, and flag specific ones, e.g., ones that have been ordered in hardcopy. ISI also has an optional bibliographic formatter with which the user can print out references that have been cited in a paper and tailored to various journals' style requirements.

The last feature – enabling the user to format citations in one of several standard reference styles – is a unique capability. As DTIC and NLM contemplate ways of providing the user with more capabilities that can suitably reflect the individual's environment, this post-processing feature of reformatting downloaded citations is worth reviewing.

8.2.3 Pro-Search

Pro-Search can automatically dial and sign onto any DIALOG or BRS database and keep track of searching charges by database, subject, and client. A Pro-Search main search screen, shown after the user has selected the "Bio & Medicine" categories and further specified "Psychology" as the subject, is shown in Figure 8-6.

Pro-Search supplies descriptions of each of the databases offered by BRS and DIALOG in a side-by-side card-file format. The user can compare categories, coverage, and cost between the two services and among the databases available. These database descriptions are updated monthly. The graphics-oriented screens use light bars and shading to guide the user through the selections.

Context-sensitive help is available, and the manual is organized in an accessible manner, complete with an index and a detailed table of contents. A function key template is also included.

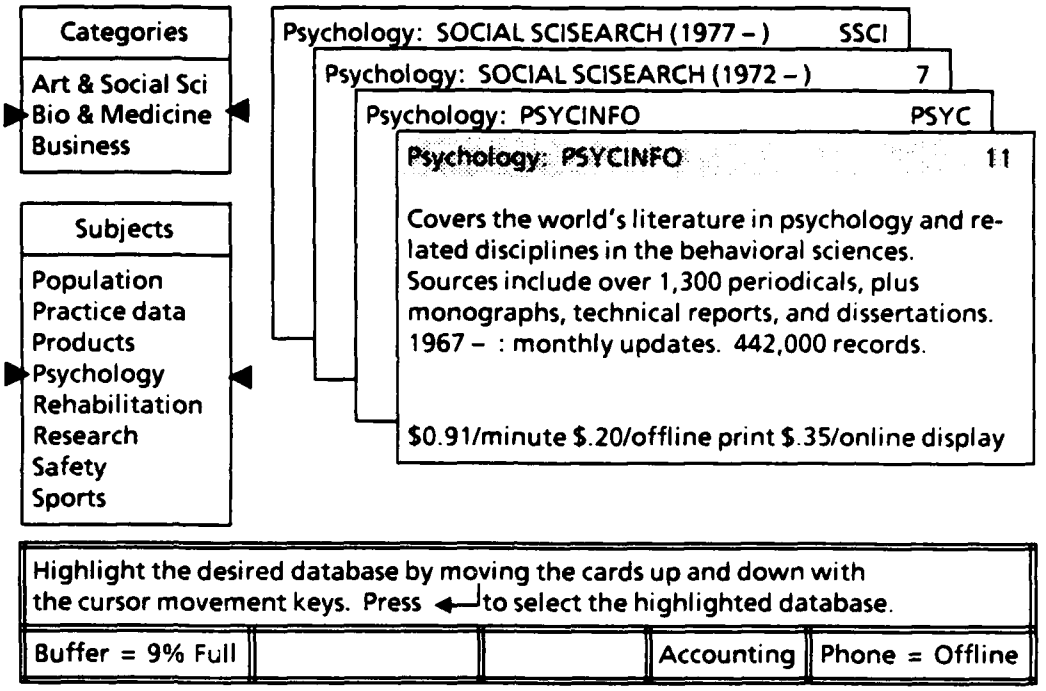


FIG. 8-6. PRO-SEARCH MAIN SEARCH SCREEN

Pro-Search has a combination of menus, on-screen prompts, and do-it-yourself features. The user can enter queries one at a time, then retrieve them by pointing to them with the highlight bar and selecting "Retrieve" from a Lotus 1-2-3-like menu. Users enter search specifications on individual lines, then combine them with standard Boolean logic. Tested search strategies can be saved on a disk for later use.

On-screen prompts enable the searching to occur on BRS or DIALOG without user interaction. This is called the "high-level search" mode. A separate "native" mode supports searches for any database. The user can define up to 23 database procedures.

Accounting capture and reporting software is available. Monthly files store search description, client, search number, search date, and detailed charges. A separate accounting program can print client invoices, showing which databases were searched and the cost of each search.

Additional software support from PBS can enable the users to process downloaded records with Biblio-Link, putting them into a form acceptable to the Pro-Cite

database. With Pro-Cite the user can create bibliographic entries for manuscripts in most popular formats.

The comparative review of database coverage, categories, and costs provided by BRS and DIALOG is a unique feature of Pro-Search. If users are faced with a choice of beginning a search in more than one database – which is not an unusual circumstance – such a gateway feature can permit comparative "shopping." This capability can help relieve a user of some of the frustration of being presented with what may seem to be an unwieldy list of available databases. The potential for this situation can already be observed in three of the four gateways compared in this report.

8.2.4 PC Net/Link

With PC Net/Link, from the Informatics General Corporation, although the user must know how to use each online system accessed, an online summary of the basic commands of each system can be displayed before the search begins. PC Net/Link also enables users to switch from one online service to another by pressing two keys – one to log off and the other to log onto the next system.

Obviously the summary of commands of basic searching in a chosen database can alleviate the need for a user to retain printed reference manuals or expend expensive online time when requesting a display of commands that have been logged onto a vendor or database. This learning tool approach for searching is also the foundation for the IIDA project (see Section 8.1.2).

8.3 PC ENVIRONMENT – VENDOR OR DATABASE-SPECIFIC GATEWAYS

Like Grateful MED, these PC-based gateways or interfaces have typically been developed by the database vendor in efforts to improve access to their products. Since they are structured to interact only with a select population of databases, these gateways are potentially harder to market. These gateways are more specialized, however, because they are usually designed by the organization that created and maintains the databases that are searched by the gateway. These products can, therefore, be quite successful in guiding the user to design a search strategy that results in greater recall and precision.

8.3.1 Wilsearch

The H.W. Wilson Company, which produces 23 secondary source publications, including *Readers' Guide* and the *Applied Science & Technology Index*, designed Wilsearch to meet two major objectives: (1) to simplify online searching to a degree where a library patron could conduct online searches without any previous training, and (2) make online searching affordable to the library patron by shifting the bulk of the processing from the more expensive mainframe to the microcomputer. Wilsearch is a menu-driven front-end system. Selection from all 23 of the WILSONLINE databases can be made by the search package, based on broad search categories identified by the user. With Wilsearch, the user can search from 43 access points, including author, title, subject, organization, journal title, date or range of dates, and classification number. It is possible to do Boolean searching (AND and OR), to use truncation symbols, and to repeat the same search automatically on different WILSONLINE databases. Updated twice weekly, WILSONLINE now offers more than 2.5 million records and is growing at the rate of half a million records a year. WILSONLINE is also available through Telebase Systems' EasyNet.

When the user has developed a search strategy by completing the search screen, the package automatically logs on, searches, shows the results, and downloads the records to the disk on the microcomputer. Wilsearch has the capability to link up to eight files for simultaneous searching. This capability is similar to DIALOG's OneSearch searching enhancement, except that here all the databases are under the control of one company (and, in this case, database vendor). Search strategies can also be saved for later use. After downloading, the system automatically logs off, and the user then has the opportunity to review the downloaded records on the microcomputer and determine which ones to print.

8.3.2 MicroDisclosure

The MicroDisclosure software from Disclosure Information Group, Inc. provides online access and downloading capabilities for the Disclosure Online Database in a menu-driven environment. This database contains annual and quarterly income statements, balance sheets, 5-year summaries, stock ownership, and other specific information derived from the reports filed with the Securities and Exchange Commission, the New York Stock Exchange, and the American Stock Exchange. Software features include searching and screening on 85 categories and

items. Reports that are preformatted, custom-generated, or both can be generated. Data files can be converted into Document Interchange Format (DIF) for further use in Visicalc, Lotus 1-2-3, and other software packages.

This post-formatting capability for numeric and/or data files, similar to Sci-Mate's ability to reformat textual records, may be worth future consideration by both DTIC and NLM. Both will need to explore how to treat records comprising numbers or raw data, since access to such information is increasingly being requested by users.

8.3.3 Online Access to Knowledge

The OAK system, designed by Charles Meadow who was also involved in the IIDA project (see Section 8.1.2), consists of tutorials and user-assistance software that enable users to search online databases at the U.S. Department of Energy's Technical Information Center. The first operational version of OAK was made available to DoE users in the spring of 1986.

OAK combines menus with other conversational methods so that users need not learn a new command language. OAK offers two computer-assisted instruction courses, one covering database searching in a broad introduction, the second telling how to search with OAK. Search assistance is provided with initial query formulation, browsing and analysis of retrieved records, and in deciding how to modify an earlier search formulation. A basic design principle has been that the user maintains control over the search at all times but may request suggestions from OAK. The software is designed for an IBM PC or PC-compatible microcomputer with a modem.

Much like DGIS or Pro-Search, OAK is not designed to be a single, general program to fit all users and retrieve from all databases. The OAK system has two major components: tutorials and assistance programs. Tutorials are for instruction before the user begins the search. Assistance programs help in the actual search.

The principal objective of the tutorials is to establish in the user's mind a mental image of the retrieval system and what the user should expect when employing the gateway. The tutorials do not teach commands but, rather, functions and concepts, such as: what is a database and a database search system.

The assistance programs provide for teaching as well. The designers determined that some end-users will attempt to search without using the tutorials or reading the users' guides. The assistance programs provide extensive help options so that a user who wishes to be instructed, step by step, can receive such a level of guidance, and an experienced user can work more rapidly without the encumbrance of elementary questions.

There are two primary ways that a user can conduct a search using OAK. One is for a user to compose a search strategy. The other approach, as with Micro-CSIN, enables the user to use a prepared search strategy, or script, available from OAK. In the latter approach, the user need only provide values of some specific search terms. A user who frequently needs to conduct a search on the same topic can create and maintain a small library of script searches for continual use.

A user constructs an OAK search by first establishing a natural language statement of the search objective. This approach is similar to the one used by CONIT (see Section 8.1.1). OAK then subdivides the main topic into facets or categories, such as subject, author, language, and date. The user is not asked to define these facets into Boolean logic. Next, each facet or category generates a list of terms that will be used in the search. The user then selects the terms, with the assistance of OAK, and the search strategy form is completed. The Boolean logic that is implied in the strategy is that all terms within a facet are connected with the OR operator and that facets are treated as sets of terms and are connected by the AND operator.

The user can review, compare, and revise the search strategy once it is completed. When the user has approved the strategy, it is automatically translated into the command language of the DoE database and communicated to the database. These translation and communication steps are not seen by the user.

Once search results are submitted to the user, judgments about whether the results are relevant must be made. Users can measure the relevancy or value of retrieved records by any suitable criteria. The user can then decide whether the search strategy should be edited, whether a new search strategy should be developed, whether review of the retrieved records should be continued, or whether the session should be ended.

Similar in function to IIDA, OAK provides a tutorial for users who will be searching a finite set of identifiable databases. For NLM, such techniques may be

beneficial when contemplating enhancements to Micro-CSIN (for scripts that do not include BIBLIO) and Grateful MED, which search a particular set of biotechnology databases.

SECTION 9

REVIEW OF GATEWAY RESEARCH

In addition to the continuing gateway design, development, and enhancement work by both DTIC and NLM, efforts in academia and private industry have also been in progress for about the past decade. Some of these efforts are described in Section 8. Here we examine some of the forward-thinking research that has been reported in the professional literature and at various conferences. This review is not designed to be exhaustive. Rather, it looks at some design approaches that have been considered on behalf of DTIC and NLM and complements the comparative analysis and review of other gateway packages provided in Section 8.

The approaches that have been drawn from the literature are presented in the context of the major components that make up a gateway, as defined in Section 1.3. Recently, however, more expanded definitions of the gateway components concerned with presearch strategies and search execution have been introduced by some researchers in the field. According to some in the field, these two additional components, if fully developed, should consist of the following functional pieces:

PRESEARCH

Problem state	Determine position of the user in the problem identification process.
Problem mode	Determine appropriate mechanism capability.
User model	Generate description of user type, goals, knowledge, background, and experience.
Problem description	Generate description of problem type, structure, topic, and context.

SEARCH EXECUTION

Dialogue mode	Determine appropriate dialogue mode for specific situation.
Retrieval strategy	Select terms and database. Build search strategy. Specify retrieval technique.
Response generator	Determine propositional structure of response, limits, and constraints to response.
Input analyst	Convert input from user into structures appropriate for other experts.

For future gateway development and enhancement, this more detailed allocation of functions among these two gateway components may be appropriate for both DTIC and NLM.

One major new approach to gateway design is an "object-oriented" gateway. It has three goals: preserve the underlying system autonomy, be implemented as a "mapping" or "translation" service, and continue to evolve as more data resources are identified. The idea is that a gateway should be capable of identifying and locating information and data – whether bibliographic, numeric, or graphic – and present it to the user in a totally transparent manner. Every "object" must be fully predefined to a level as specific as possible so that the gateway can relate it to the user's request. For bibliographic information, the lowest level may be the words in the citations or full-text documents and their syntactic relationships. For graphics it may be each of the individual components, fully identified in a database, that make up a large engineering drawing. Several prototypes that incorporate conventional DBMSs are in development and could be of particular interest to both DTIC and NLM as each organization continues to expand its resource offerings to its users.

9.1 CONDUCTING PRESEARCH ACTIVITIES AND SELECTING DATABASES

This component is perhaps the newest area of research in gateway development. Search strategy formulation and refinement is a sophisticated feature not widely available on gateways but very much needed. AI is being applied to handle the interaction required between a user and a database, as demonstrated by projects such as CONIT. But much needs to be developed in this area. Gateways need to

assist the user in formulation and refinement of search strategy – steps essential to a useful search.

In a traditional intermediary environment, search terms and databases are not selected without an understanding of the user's situation. A human intermediary generally does not ask the user to specify a search topic but tries first to gain an understanding of the context in which any eventual search topic will be embedded. To some authors, it seems reasonable that gateway designers should not shift the burden onto the user for successfully translating such an understanding of a user's information request into language or subject terms appropriate for database submission. Therefore, they suggest that an intelligent gateway must at least incorporate these model-building functions and response-generating functions and must operate within a highly interactive, cooperative dialogue with the user. Some approaches for allowing a gateway to gain an understanding of the user's information request without a human intermediary and to provide enhancement search strategy and database selection capabilities are described below.

9.1.1 User-Level Assessment Aid

Martha Williams, in her 1986 article titled "Transparent Information Systems Through Gateways, Front Ends, Intermediaries, and Interfaces" published in the *Journal of the American Society for Information Science (JASIS)*, suggested a "user-level assessment aid," which would determine automatically whether the user was a scientist, student, or manager by looking at the content and form of the query and applying stored data concerning the user. This capability would require AI techniques that now exceed the state of the art. However, it could build on the work accomplished by the gateway developed by DIALOG, which presents menus to a user depending on the type of searching needs and experience of the user.

9.1.2 Tailoring Gateway Interaction to Individual Users

During the 1986 American Society for Information Science (ASIS) Annual Meeting, Christine Borgman reported on research during her session "Incorporating Users' Information Seeking Styles into the Design of an Information Retrieval Interface," comparing individual user indicators that would point to an aptitude for success with an online retrieval system.

Dr. Borgman's hypothesis is that, for as useful as gateways can be to both end-users and search intermediaries, each is designed as though all users were the same. No attempt is made to tailor the interface to a particular user. Dr. Borgman's understanding of recent research had shown her that users perceive systems differently from the way in which their syntax, database structure, and content are formally specified. Users' perceptions vary by previous experience and training, subject specialty, age, and cognitive skills. Though the effects of training and experience are more widely known, subject background and cognitive style are less well-researched although it is suspected that they have an effect on use of retrieval systems.

This approach was further underscored by Dr. Trudi Bellardo[1] who defined online literature searching as a series of interrelated steps and complex tasks. These steps and tasks are reflected in the functional pieces which have been allocated among the presearch and search execution gateway components on page 9-1 and page 9-2. The hypothesis is that individuals possess different critical and analytical thinking approaches, and that even with similar educational and database searching instruction backgrounds, may approach search strategies from different perspectives and consequently, with different retrieval results.

We can conclude from these lines of research that gateways need to relate uniquely to every user and to that user's perception of the database and the retrieval methodologies.

9.1.3 Searchers Versus Indexers: Different Thesauri

Showing searchers a network of subject terms and term relationships can suggest ideas for browsing and help users decide what terms are appropriate for their searches. Such a capability can also suggest additional terms that may be needed to capture the full scope of terminology that describes their query.

Existing thesauri are designed primarily for indexers and only secondarily, if at all, for intermediaries and end-users. Such theasuri enable indexers to identify "legitimate," or permissible, terms with which to index documents. The searcher must, therefore, use the same terms (if searching on controlled vocabulary). Marcia

---

[1]Bellardo, Trudi. "An Investigation of Online Searcher Traits and Their Relationship to Search Outcome." *Journal of the American Society for Information Science* 36/4 (1985): pp. 241 – 250.

Bates in her 1986 *JASIS* article "Subject Access in Online Catalogs: A Design Model" reiterated this need.

Ideally, a gateway should support an end-user thesaurus that would list all the terms used to index a document and use self-explanatory codes and cross-references. The variety of initial entry terms a searcher may use is enormous. An effective end-user thesaurus should, therefore, comprise a vast number of entry terms, which can then lead the user to legitimate terms. The searcher should be able to enter any term, legitimate or not, and be shown the conceptually closest legitimate term, as well as related descriptors to use in searching.

9.1.4 Navigational Aids

Even within a hierarchical gateway framework that has clearly defined menu and/or command trees, the sheer volume of information makes it easy for a user to become lost. Consequently, development of navigational aids should be considered. There are three types: (1) passive – where color or sign posts indicate the information level, (2) video – where a second screen display depicts a world view of the system with a simultaneously displayed screen that provides the detailed or local view, and (3) audio – where icons emit a sound associated with a specific information level. Alternate formatting through color, flashing, flagging, magnification, space, and time can improve a display without changing the information content of the screen. In a system that is capable of windowing and clustering the displayed information, such a feature can enable a user to retain one part of the display on the screen (without having to remember it) while displaying new information in another window. Such windowing techniques and their utility were documented by R. G. Crawford and M. S. Becker who implemented the Friendly Information Retrieval SysTem USERS' (FIRSTUSER) interface which was integrated to INGRES and accessed a bibliographic database.

9.1.5 Layout Aids

"Layout" refers to the ability of a system to dynamically provide the appropriate display, based on the information or data types that satisfy a query. The display must be modifiable by the user. It should also be capable of incorporating both updated and more detailed information. The technical solution to developing layout aids relies on knowledge-based graphics and real-time image generation.

Graphics display is dependent upon the descriptive information available and the rules resident in the database.

9.1.6 Teaching Online Retrieval Skills

An online searching instructor (either human or machine-aided) can either teach the use of a particular retrieval language and then move inductively toward general principles by examples drawn from that language, or present general information retrieval principles, deduce examples that are not tied to any one system, and then apply the principles to a specific system. Work in this area has been done by the IIDA, OAK, and I^{3}R projects.

9.2 EXECUTING THE SEARCH STRATEGY

Automated dial-up and log-on to a telecommunication network or vendor is the most developed of the gateway components. Enhancements in this gateway component have focused on improved feedback to the user during the search execution period so that a search strategy may be improved either before or while logged onto a database(s). Some of the suggested improvements that have been discussed are presented below.

9.2.1 Automating the Search Refinement Process

Martha Williams, in the 1986 *JASIS* article "Transparent Information Systems Through Gateways, Front Ends, Intermediaries, and Interfaces," suggested an "automated refinement" of the search strategy component. Automated refinement of a search can employ such techniques as pattern matching, clustering, and other associative devices that may be used in the presearch phase, during or after the search. These techniques can operate on virtually any data element that is in the database being searched. Presearch analysis could be done for user profile refinement, in-search analysis to help direct the search, and post-search analysis for the purposes of sorting out the better material or ranking output on the basis of objective criteria. Such refinements, as suggested by Williams, could be particularly useful when applied to the object-oriented gateway environment.

Additionally, many researchers (N. J. Belkin particularly) have argued that it is fallacious to assume that an individual with an information need approaches a database with a reasonably well articulated query. To state their needs, users may have to describe what they do not know. In effect, users may not have a "query"

per se, but an "anomalous state of knowledge." Traditional approaches of matching an incoherent query against a database that has represented documents systematically in a coherent manner may be forcing a comparison between two dissimilar frameworks. By permitting a user to continue to refine the search process from an initial visceral question to a more formalized need can greatly increase the effectiveness of a search-and-retrieval session.

9.2.2 Routing User Query from "Meta-Directory" to Specific Source

Search strategy formulation, although acknowledged as quite difficult for end-users, can be further compounded by the need to decide which database(s) are suitable. As reported by Geraldene Walker in her 1986 article "End-User Searching: The Beginning or the End?" in *The Reference Librarian*, "a database selection [by end-users] was their biggest problem." Several studies have shown that an end-user needs more time than a trained intermediary to conduct a search and choose appropriate database(s). Therefore, identifying the database(s) and getting the user's search strategy submitted successfully to those databases is a critical aspect of providing for a successful gateway.

A directory of directories, such as the one now under development by DTIC, could be functionally expanded so that it not only assisted the user in identifying the source(s) or database(s) most pertinent to the information query, but automatically routed the query to the target database(s). Micro-CSIN also helps the user preselect databases for search strategy submissions, but provides the user with only minimal database background information.

Research suggests that selection of appropriate sources for the user could not only include the subject content available from the database(s) but also review some or all of the following:

- Vendor selection based on database availability, charges, size of current and back files, file updating schedules, user familiarity with the vendor, gateway capability to use a CCL to assist the user in accessing an unknown database with a known retrieval language, and auxiliary services such as document delivery. The Pro-Search PC-based gateway now provides a basic comparative review of available databases for users.
- Telecommunications network selection based on charges, location, reliability, availability, and other user-defined needs.

- An initial approach to the database would be selected on the basis of whether a user was more comfortable with English-like natural language query structures or was familiar enough with the controlled vocabulary used by the database. Support from the gateway – for example, an end-user thesaurus (see Section 9.1.3) – would be included in this selection process.

9.3 PERFORMING POST-PROCESSING ON THE RESULTS OF A SEARCH SESSION

As a few packages can now do, gateway software should be expandable to incorporate additional applications software. For example, after completing searches in one or more of the databases, the search output might automatically call up the appropriate application package for further manipulation or formatting of the output. Application packages would include report generators, word processors, statistical packages, and econometric models. They should be able not only to handle retrieved data but also to incorporate user data, which could be merged or uploaded for processing. In some cases, these application packages might be downloaded to the user site for local processing.

9.4 GATEWAY FUNCTIONALITY

Use of the UNIX operating system as an IGP was the approach taken by LLNL when developing the TIS, which then migrated to DTIC as DGIS. UNIX will also be used to develop the next major releases of both Micro-CSIN and Grateful MED. The availability of shells, screen packages, window packages, and menu prototype systems, as well as the ability to support a wide set of device drivers, have made UNIX an excellent gateway development operating system. The first IGP connections were made to bibliographic and numeric online database systems, demonstrating for the first time the ability of UNIX to connect heterogeneous systems. Organizations that realize they must acquire new systems or expand existing ones, find that a UNIX foundation helps to prevent incompatibility situations; the disparate systems can connect to one another via an IGP.

UNIX will most probably be the operating system of choice as organizations wish to exchange graphics and voice message over existing local and long-haul networks. Since vendors that supply the equipment for these functions usually do not provide interconnectivity with other systems, use of IGPs based on UNIX may prove to be appropriate in the future, as gateways continue to encompass more data

types. This may be important as organizations in the future wish to adapt the object-oriented gateway approach described at the beginning of this section.

Another important change that will affect gateways that are designed primarily for PCs is the recent introduction of OS/2. Most PC vendors expect to place OS/2 on all 286 and 386 computers and to offer this new operating system as the only option. Market surveys suggest that OS/2 will supersede DOS as the PC operating system of choice by the end of 1990. This schedule will be accelerated to mid-1989 for business PC users. Both NLM and DTIC should assess use of this PC-based multi-tasking operating system when considering new versions of PC-based gateways.

SECTION 10

RECOMMENDATIONS

As discussed in prior chapters, the fundamental capabilities of a gateway can be categorized as features, utilities, support for users, and functionality. A successful gateway should provide the user with the following capabilities for each of the four basic capabilities:

- Features including presearch strategy development, database selection, search execution, and post-processing should provide:
  - Assistance in structuring a search strategy with an online reference to controlled vocabularies.
  - Straightforward use of multiple database retrieval languages by use of prompts, hints, and assistance with such means as query analysis, translations or CCL, menus, and expanded prompts so that searches for one database can be applied to another and terms retrieved from one database can be turned into queries for another.
  - Guidance and access to appropriate databases with support information presented using explanations and examples.
  - Iterative feedback during a search session so that searches can be modified to closely represent the user's information need.
  - Post-processing of the information retrieved as the result of searches conducted in multiple databases to include the elimination of duplicate citations, as well as the merging and sorting of citations. Suggestions for improving the initial search strategy, e.g., additional relevant subject terms, should be offered.
- Utilities should include:
  - Facilitated file management so that users can name, edit, modify, delete and perform other needed functions to handle the files created as the result of gateway use.
  - Available subsidiary accounting functions that permit users to discriminate the costs incurred by use of the gateway.

- Support for users should include:
  - An interactive help subsystem that presents messages in a context that is meaningful to the experience level of the user.
  - Online tutorials that present, for example, retrieval skills, editing features, and file management functions that complement the online help subsystem.
- Functionality, for example, the operating system and use of menus should provide:
  - A foundation that is based on a versatile and portable operating system.
  - Consistent presentation of menus and commands invoked by the users.

These fundamental gateway capabilities form the basis for our recommendations for DTIC, for NLM, and for overall direction for gateway developments for bibliographic information users.

10.1 RECOMMENDATIONS FOR DTIC

The gateways supporting DTIC should be enhanced to meet the fundamental capabilities described above. The following changes should be made:

- DGIS currently offers no presearch capabilities. No guidance is provided on how to construct a search strategy except for the examples provided in the User's Guide. Even with the addition of the CCL and the online Directory of Resources, a major gap in assisting the end-user construct these strategies will exist. We recommend that DTIC review the accomplishments of R&D projects, such as CONIT, and existing gateways, such as Micro-CSIN, Grateful MED, Sci-Mate, and Pro-Search. These gateways lead users through a systematic identification of their search requests before any online activity occurs. Also online access and guidance to the DTIC Retrieval and Indexing Terminology (DRIT) should be provided for DROLS searching as exampled by Grateful MED's interaction with NLM's MeSH. Ultimately, a DGIS user should be able to create a natural language search query and have it submitted and translated to the vocabularies of selected databases.
- DGIS post-processing should be expanded to recommend relevant search terms to the user once a set of relevant citations has been identified by the user. Grateful MED has this capability. DTIC should initially focus on DROLS searches with search term recommendations made using the DRIT. This feedback mechanism can be particularly helpful to users who may not be familiar with how documents have been indexed for the database.

- If DTIC completes the gateway enhancements that are now in process (see Section 7.1) and implements the presearch features recommended in this report, the capabilities of DGIS to respond to end-user searching techniques are greatly expanded. If DGIS can offer both experienced and novice users search strategy and database selection guidance suitable to their experience level, then access to SearchMAESTRO is no longer warranted.

- If DTIC continues to provide access to SearchMAESTRO, then post-processing features with capabilities similar to ones now available in DGIS should be created. The transformation, merger, and processing of retrieved citations into a bibliography is of great benefit to a user – whether intermediary or novice end-user. Consistent provision of capabilities across all of the gateways or interfaces offered by DTIC will ultimately benefit the user population who may otherwise become confused and disinclined to use these gateways as a result.

The following enhancements should be made to DTIC gateway utilities, support, and functionality:

- The existing EM utilities of DGIS are helpful to the user population. Further improvements to the editing capabilities offered by either Visual Editor (Vi) or the line editor (Ex) to provide for more word processing-type functions are recommended.

- User support should include a clear path for novice end-users as to which gateway is most suitable for them. If SearchMAESTRO lets a user search in DIALOG and other databases, which are also accessible via DGIS, the user should be directed to the correct gateway or interface. We recommend that the DGIS be improved, as discussed above, and that use of Search-MAESTRO be phased out.

- DGIS must be able to guide users to appropriate data resources. An understanding of the "user model" as handled by the I^{3}R experimental gateway (see Section 8.1.4) should be reviewed and the concept built into the online Directory of Resources. If the gateway is better equipped to identify the user's perception of the information request (or user "model") then better direction to relevant resources can be implemented.

- Functionally, the UNIX foundation for the current DGIS is technically sound. The decision DTIC may need to face in the future is which operating system to use – UNIX or O/S 2 – for any PC-based gateway software that may be issued.

- There are inconsistencies in DGIS in how some commands are used, e.g., "q" to leave EM but "e" to leave other comparable gateway functions. These are minor problems. Overall there is easy maneuverability between the DGIS menus and commands.

10.2 RECOMMENDATIONS FOR NLM

Basically both NLM gateways effectively perform the features of presearch strategy development, database selection, search execution, and post-processing. Recommendations on how to enhance the features of these gateways are provided in Section 10.3.

The following enhancements should be made to NLM gateway utilities, support, and functionality:

- Grateful MED does not retain any search results, beyond the temporary file of records retrieved by a search, available for initial review by the user. Once the user has reviewed and printed out the citations, the file is replaced the next time the user conducts a search. Enhancing the Grateful MED package with Micro-CSIN post-processing features has eliminated this problem. However, if NLM continues to offer Grateful MED as a stand-alone package, this file handling problem must be addressed.

- Micro-CSIN should be more user-friendly. The menu-approach capabilities of SearchMAESTRO and help screens provided by Grateful MED are good examples of how the user can be more easily led through the structured search strategy development, execution, and post-processing routines. Although these latter two gateways have been designed primarily for end-users, whereas Micro-CSIN supports intermediaries, these gateways have relevant concepts for user interaction which can benefit Micro-CSIN. Both Micro-CSIN and Grateful MED should be designed to present help messages in a window shown simultaneously with the selected functional screen.

- Functionally, NLM needs to determine how to view its gateway products – Micro-CSIN and Grateful MED – with respect to the overall gateway planning that has begun as the ABIDE framework. As now faced by DTIC, NLM will be able to offer its users a selection of gateways to gain access to both in-house NLM and external information resources. These currently available PC-based software packages may duplicate the ABIDE gateway functions. Consistent access by users and overall gateway expansion planning needs to be determined now. The experience of DTIC in developing the online Directory of Resources may be of benefit to NLM to address this recommendation.

- NLM needs to choose where the majority of gateway processing is to occur – either on a minicomputer/central server or in a PC-based/ distributed environment. If both environments are to be supported, then NLM must carefully plan how functions can be developed in parallel in both environments and recognize the distribution problems of PC-based software and expanding support problems incurred by the centralized approach.

10.3 OVERALL RECOMMENDATIONS

Emphasis should be placed on enhancing gateway presearch features with continued use of expert systems and AI. This recommendation is comprised of the following parts:

- Both DTIC and NLM should review the "user" and "request" model approach developed by the I3R gateway project. The purpose of these models is to produce a session of retrieval activities that is similar to a human intermediary.
- Improved means for structuring search strategies tailored to a user's knowledge of the subject area and online retrieval techniques are recommended. The CONIT and I3R gateway projects and DIALOGLINK product all offer experience with developing some of these capabilities.
- Users, after submitting a natural language query, should be provided with the ability to review and use terms from diverse thesauri without needing to know the controlled vocabulary of any database. The CODER and OAK gateway projects, as well as Dr. Marcia Bates' research, have explored this capability.
- Presentation of relevant data resources, based on the user's search strategy, should be available. Users should be able to choose among possible resources based on such criteria as subject coverage, online charges, depth of coverage, and update frequency. Once selected, the gateway should automatically log onto the selected database(s) and execute the search strategy.

Presearch functions are the most critical aspect of database searching. A gateway, no matter how elegant its interfaces to numerous heterogeneous databases, cannot compensate for an ill-conceived search strategy that does not address the user's true information need.

The following enhancements should be made to both DTIC and NLM gateway utilities, user support and functional capabilities:

- Improved accounting utilities should be provided so users can easily track online usage and charges. The Pro-Search package offers one example for review.
- Online assistance should be in context and support the user in understanding how to best use the commands and menus. Messages that are meaningful, such as those devised in the IIDA experiment, are recommended. IIDA messages related procedural or syntactic errors to the

user in a nonthreatening manner so that "fatal error" messages were replaced with "You have not defined set 4 yet".

- New releases of any of the DTIC or NLM gateways should be designed using operating systems and software packages that allow for versatility and portability. Both DTIC and NLM should review the object-oriented approach to gateway development. This approach, which supports retrieval from databases of dissimilar contents, will become increasingly important as DTIC and NLM consider ways to access all resource and database types – bibliographic, numeric, and graphic.

As DTIC and NLM provide access to more nonbibliographic databases, attention must be paid as to how to successfully download and post-process such data. Migration to an object-oriented gateway structure may assist both organizations in bringing together these diverse resources for their users.

Since both DTIC and NLM are moving to broader environments, the potential for resource sharing and cooperative development is propitious. We recommend that these two organizations consider ways to initiate complementary gateway enhancement efforts.

10.4 GATEWAY DEVELOPMENT AND ENHANCEMENT ISSUES

The following sections highlight some issues that need to be addressed by one or both organizations before gateway development is continued.

10.4.1 Use of a Gateway by an End-User

Use of a gateway by a novice database searcher (or end-user) raises several issues. An intelligent (and user-friendly) gateway can aid the novice searcher, as well as the casual or occasional searcher, who does not wish to memorize the detailed specifications required for online searching for the multitude of databases applicable to a user's information need.

On the other hand, gateways can remove the end-user from the interactive aspect of the development, execution, and iterative modification of a search strategy. Less interaction can lead to reduced search strategy effectiveness – particularly in precision and recall. Some gateways have selected a simplistic approach to assisting end-users retrieve information from complex databases. This approach, illustrated by SearchMAESTRO, retrieves a limited number of recent citations. This cannot be considered as comprehensive as searches performed with a more sophisticated

gateway, such as Micro-CSIN. As has been frequently documented, scientists and engineers have "tended to need exhaustive searches in specific subject areas . . . to be used in support of ongoing research projects, or in preparing proposals for future research."[1] This need for exhaustive searching can be inadvertently defeated by the retrieval approach taken by several gateways reviewed in this report. With too simple a gateway, end-users may miss important citations, and because of their ignorance of database coverage and retrieval skills, they may mistakenly assume that they have reviewed all the relevant literature. It is important for both DTIC and NLM to consider such end-user information-seeking behavior when planning how to extend gateway capabilities.

10.4.2 Gateway Hardware Environment

As described earlier, there are two basic gateway hardware environments: mainframe/minicomputer (centralized) or microcomputer (decentralized). Each environment has advantages and disadvantages. As DTIC and NLM plan versions of their currently available gateways for other environments and/or as they develop a new gateway (for example, an intelligent workstation), the following issues should be kept in mind.

10.4.2.1 Distributing and Updating Gateway Software

Either gateway location – centralized (usually mainframe/minicomputer) or decentralized (typically PCs) – present both distribution and ongoing update problems. Gateway software located on a centralized resource presents the fewest distribution difficulties. Once the software is placed on and made available to users from the mainframe/minicomputer, then distribution, with the exception of disseminating user documentation, has been completed. Initial distribution of PC-based gateway software is not difficult either, but accurate user records must be maintained to facilitate the distribution of subsequent updates.

As enhancements are added to a gateway, then an update is warranted. Updates may also be needed when database vendors modify access or search procedures to their databases rendering the existing gateway obsolete until an updated version is created and distributed. If the gateway is located on a central

---

[1]Case, Donald, Christine L. Borgman, and Charles T. Meadow. "End-User Information-Seeking in the Energy Field: Implications for End-User Access to DOE/RECON Databases." *Information Processing & Management* 22/4 (1986): pp. 299 – 308.

machine, software updates and enhancements are readily provided to the user community by updating the software on the central computer. Gateway software in a PC-environment removes the processing demand placed on centralized computing resources to support gateway functions for all users, but creates distribution problems. When PC-based software has been updated, it must be distributed to each member of the user community. Unless accurate records exist, some users may not receive updated software. If practicable, updates to PC-based gateway software can be downloaded to users via telecommunications, but such procedures require identification of users who can accommodate these measures.

10.4.2.2 Accessing Databases with Gateway Software

Telecommunications to the databases available from the gateway are most easily supported by multiple modems or modem pools associated with a central processor. As gateways evolve and are able to provide users with the opportunity to simultaneously search several databases, multiple modems are a necessity. PC-based software, particularly that using DOS, cannot support such multiple processing. Modem pools on LANs that support PCs typically are not available and, therefore, eliminate the capability to support simultaneous multiple database processing. Perhaps as PC software is created under newer multi-tasking operating systems, such as O/S 2, some of these problems will be alleviated.

A centralized gateway can support a broad range of existing dumb terminals, PCs, and other devices capable of telecommunications. Commonly-available telecommunications software resident on a mainframe/minicomputer can provide a user with access to the centralized gateway that permits searching in a large number of databases. However, the additional layer of telecommunications may limit searching to lower-speed asynchronous baud rates and create slower response times than if the user were directly connected into the database. In addition, increased demand for access to the gateway on a centralized machine can quickly outstrip the available resources on that machine.

10.4.3 Impact of New Storage Technologies

Other technologies, such as advances in local storage media as exemplified by CD-ROM, can place the contents of a database at the local level for searching and retrieval. Additionally the availability of local intelligence to support gateway software, such as intelligent terminals or workstations, can also allow effective

gateway activity at a local level. This approach contrasts with the current approach now provided by the four gateways available from DTIC and NLM. Each approach has advantages and disadvantages.

If a database is searched on a CD-ROM, the costs associated with maintaining a database can no longer be shared with other organizations that once accessed the database on an online basis. Instead, the library or information center must provide the CD-ROM equipment and track all subsequent updates to the database. These additional costs may be reflected in user chargeback fees. If use of CD-ROM expands, smaller online databases may disappear more rapidly since demand for them in online form will diminish more rapidly. This decreased online demand may be interpreted as meaning there is no need to convert these databases to an alternate storage technology and they become defunct. Conversely, these technologies are attractive since they offer high-density recording and low-cost duplication for distribution of already-mastered discs. Also CD-ROM searchers do not have to contend with telecommunications problems or slow response times of vendor systems.

10.4.4 Privacy and Ownership

Every member of the information query chain, from user to author of the original material, is concerned with privacy. Users do not wish to reveal the nature of their search questions. Corporations do not wish to identify – by the nature of their information queries or new submissions to a database – new products, sensitive issues, market approaches or the like to their competitors. Recent evidence of this situation was seen in the 23 November 1987 issue of *InfoWorld* where it was reported that DIALOG has permitted Dun & Bradstreet to restrict access to some of their databases to a selected group of users – in this instance labor union libraries and information centers. Although this decision has been subsequently reversed, the implications for the future are obvious.

The question may become whose customer is the end-user – the database producer, the database vendor, or the gateway that provides access to the database. Providers of each level of service will have to remain aware of the potential for the invasion of privacy and the ownership issues, resulting in "information embargoes."

GLOSSARY

ABIDE	=	Advanced, Biomedical Information Databand Exchange
AG	=	application gateway
AI	=	artificial intelligence
AMA	=	American Medical Association
AMANET	=	AMA Network
ARPANET	=	Advanced Research Projects Agency Network
ASCL	=	Advanced Super Computer Laboratory
ASIS	=	American Society for Information Science
ATSDR	=	Agency for Toxic Substances and Disease Registry
AT&T	=	American Telephone and Telegraph
BIBLIO	=	Bibliographic
BRS	=	Bibliographic Retrieval Service
CAS	=	Chemical Abstracts Service
CATLINE	=	CATaloging OnLINE
CCL	=	common command language
CCLS	=	CCL System
CDC	=	Center for Disease Control
CD-ROM	=	compact disc read-only memory
CHEMID	=	Chemical Identification
CIS	=	Chemical Information System
CODER	=	Composite Document Expert/extended/effective Retrieval
CONIT	=	Connector for Networked Information Transfer
COPS	=	Command Pattern Search System

CSIN	=	Chemical Substances Information Network
DBMS	=	database management system
DEC	=	Digital Equipment Corporation
DGIS	=	Defense Gateway Information System
DIF	=	Document Interchange Format
DoE	=	Department of Energy
DOS	=	Disk Operating System
DRIT	=	DTIC (Defense Technical Information Center) Retrieval and Indexing Terminology
DROLS	=	Defense RDT&E Online System
DTIC	=	Defense Technical Information Center
EM	=	electronic mail
EPA	=	Environmental Protection Agency
Ex	=	line editor
FEDRIP	=	Federal Research in Progress
FIRSTUSER	=	Friendly Information Retrieval SysTem USERS
GUSTO	=	Gateway User Support and Training Office
HSDB	=	Hazardous Substances Data Bank
IAC	=	Information Analysis Center
IBM	=	International Business Machines Corporation
IGP	=	Intelligent Gateway Processor
IIDA	=	Individualized Instruction for Data Access
I^3R	=	Intelligent Intermediary for Information Retrieval
INGRES	=	Interactive Graphic and Reporting System
ISI	=	Institute for Scientific Information
JASIS	=	Journal of the American Society for Information Science
LAN	=	local area network

LISP	=	LISt Processing
LLNL	=	Lawrence Livermore National Laboratory
MATRIS	=	Manpower and Training Research Information System
MEDLARS	=	Medical Literature Analysis and Retrieval System
MEDLINE	=	MEDLARS OnLINE
MeSH	=	Medical Subject Headings
Micro-CSIN	=	Micro-Chemical Substances Information Network
MIT	=	Massachusetts Institute of Technology
NASA/RECON	=	National Aeronautics and Space Administration's Remote Console
NFAIS	=	National Federation of Abstracting and Information Services
NIH	=	National Institutes of Health
NISO	=	National Information Standards Organization
NLM	=	National Library of Medicine
NTIS	=	National Technical Information Service
OAK	=	Online Access to Knowledge
OHMTADS	=	Oil and Hazardous Materials Technical Assistance Data System
OHS	=	Occupational Health Services
PC	=	personal computer
PROLOG	=	PROgramming LOGic
RAM	=	random access memory
RDT&E	=	Research, Development, Test, and Evaluation
RECON	=	Remote Console
SANSS	=	Structure and Nomenclature Search System
SDC	=	Systems Development Corporation

SearchMAESTRO	=	Search Menu-Aided Easy Searching Through Relevant Options
SOS	=	Save Our Search
STI	=	scientific and technical information
STN	=	Scientific and Technical (Information) Network
TIS	=	Technology Information System
TOXCHEM	=	Toxic Chemical
TOXNET	=	Toxicology Network
TR	=	Technical Reports
WUIS	=	Work Unit Information System
Vi	=	Visual Editor

BIBLIOGRAPHY

Alberico, R. "Front End Games." *Small Computers in Libraries* (Jan 1986): pp. 10 – 15.

Badgett, T. "Search Software: Directory Assistance." *PC Magazine* (May 1987): pp. 263 – 273.

Badgett, T. "Tapping into On-Line Knowledge." *PC Magazine* (May 1987): pp. 237 – 258.

Bates, Marcia J. "Subject Access in Online Catalogs: A Design Model." *Journal of the American Society for Information Science* 37/6 (1986): pp. 357 – 376.

Belkin, N. J., et al. "ASK for Information Retrieval: Part I. Background and Theory." *Journal of Documentation* 38/2 (1982): pp. 61 – 71.

Bellardo, Trudi. "An Investigation of Online Searcher Traits and Their Relationship to Search Outcome." *Journal of the American Society for Information Science* 36/4 (1985): pp. 241 – 250.

Boyce, Bert R. "Computer-Assisted Instruction for Online Searchers." *Bulletin of the American Society for Information Science* (Apr/May 1987): p. 34.

Case, Donald, Christine L. Borgman, and Charles T. Meadow. "End-User Information-Seeking in the Energy Field: Implications for End-User Access to DOE/RECON Databases." *Information Processing & Management* 22/4 (1986): pp. 299 – 308.

Crawford, R. G., and H. S. Becker. "A Novice User's Interface to Information Retrieval Systems." *Information Processing & Management* 22/4 (1986): pp. 287 – 298.

Croft, W. B., and R. H. Thompson. "I^{3}R: A New Approach to the Design of Document Retrieval Systems." *Journal of the American Society for Information Science* 38/6 (Nov 1987): pp. 389 – 404.

Fenichel, Carol Hansen. "Online Searching: Measures that Discriminate Among Users with Different Types of Experiences." *Journal of the American Society for Information Science* 32/1 (Jan 1981): pp. 23 – 32.

Freedman, D. "Perspectives: Information Gateways Make On-Line Databases Easier to Use." *High Technology* 6/1 (Jan 1986): pp. 64 – 65.

Hewes, Jeremy Joan. "Gateways to On-Line Services." *PC World* (May 1985): pp. 149 – 156.

Hildreth, C. R. "Communicating with Online Catalogs and Other Retrieval Systems: The Need for a Standard Command Language." *Library Hi Tech* 4/1 (Spring 1986): pp. 7 – 11.

Janke, Richard V. "Presearch Counseling for Client Searchers (End-Users)." *Online* (Sep 1985): pp. 13 – 26.

Kesselman, M. "Front-End/Gateway Software: Availability and Usefulness." *Library Software Review* (Mar/Apr 1985): pp. 67 – 70.

Levy, Louise R. "Common Command Language." *Bulletin of the American Society for Information Science* (Jun/Jul 1987): pp. 31 – 32.

Marcus, Richard S., and J. Francis Reintjes. "A Translating Computer Interface for End-User Operation of Heterogeneous Retrieval Systems. I. Design." *Journal of the American Society for Information Science* 32 (Jul 1981): pp. 287 – 303.

Marcus, Richard S., and J. Francis Reintjes. "A Translating Computer Interface for End-User Operation of Heterogeneous Retrieval Systems. II. Evaluations." *Journal of the American Society for Information Science* 32 (Jul 1981): pp. 304 – 317.

Marcus, Richard S. "An Experimental Comparison of the Effectiveness of Computers and Humans as Search Intermediaries." *Journal of the American Society for Information Science* 34/6 (Nov 1983): pp. 381 – 404.

Meadow, Charles T., Thomas T. Hewett, and Elizabeth S. Aversa. "A Computer Intermediary for Interactive Database Searching. I. Design." *Journal of the American Society for Information Science* 33/5 (Sep 1982): pp. 325 – 332.

Meadow, Charles T., Thomas T. Hewett, and Elizabeth S. Aversa. "A Computer Intermediary for Interactive Database Searching. II. Evaluation." *Journal of the American Society for Information Science* 33/6 (Nov 1982): pp. 357 – 364.

Monarch, Ira, and Jaime Carbonell. "CoalSORT: A Knowledge-Based Interface." *IEEE Expert* 2/1 (Spring 1987): pp. 39 – 53.

Moskowitz, R. A. "Gateways for Online Information." *Computer Decisions* 17/18 (Sep 1985): pp. 36, 40.

Rouse, William B., and Sandra H. Rouse. "Human Information Seeking and Design of Information Systems." *Information Processing & Management* 20/1-2 (1984): pp. 129 – 138.

Stout, C., and T. Marcinko. "Sci-Mate™: A Menu-Driven Universal Online Searcher and Personal Data Manager." *Online* 7/5 (Sep 1983): pp. 112 – 116.

Vickery, Alina, et al. "A Reference and Referral System Using Expert System Techniques." *The Journal of Documentation* 43/1 (Mar 1987): pp. 1 – 23.

Walker, Geraldene. "End-User Searching: The Beginning or the End?" *The Reference Librarian* 14 (Spring/Summer 1986): pp. 39 – 51.

Walton, Kenneth R., and Patricia L. Dedert. "Experiences at Exxon in Training End-Users to Search Technical Databases Online." *Online* (Sep 1983): pp. 42 – 50.

Williams, Martha E. "Electronic Databases." *Science* 228/4698 (Apr 1985): pp. 445 – 456.

Williams, Martha E. "Transparent Information Systems Through Gateways, Front Ends, Intermediaries, and Interfaces." *Journal of the American Society for Information Science* 37/4 (Jul 1986): pp. 204 – 214.

"SDC, ISI Show New Microcomputer Front Ends for Searching." *Online Database Review* 4/5 (May 1983): pp. 5 – 6.

BBN Laboratories Incorporated. *Users' Guide to DGIS.* BBN Report No. 6532. May 1987.

Bolt Beranek and Newman. *Preliminary Release: Micro-CSIN User's Guide – PC DOS Version.* Sep 1985.

Cotter, Gladys A. *An Intelligent Gateway for the Department of Defense. The Technology Information System.* AD-A133 800. Jun 1984.

Cotter, Gladys A. *The DoD Gateway Information Systems: Prototype Experience.* AD-A166 200. Apr 1986.

Hushon, J. *Survey of Commercially Available Microcomputer Gateway Software.* Prepared for Executive Office of the President, Council on Environmental Quality, Environmental Protection Agency and the National Library of Medicine. Report No. 6016. Aug 1985.

Kuhn, Allan D. *Artificial Intelligence Developments Re: DoD Gateway Information System (DGIS) and Defense Applied Information Technology Center (DAITC).* DTIC AI Foundational Series No. 2. AD-A181 101. Feb 1987.

Kuhn, Allan D., and Gladys A. Cotter. *The DoD Gateway Information System (DGIS): User Interface Design.* AD-A174 150. Sep 1986.

Online Information International, Inc. *The National Library of Medicine Grateful MED User's Guide.*

Pylyshyn, Zenon W. *Intelligent Database Interfaces: A Survey of Some Artificial Intelligence Applications.* Cognitive Science Memorandum (COGMEM) No. 17. Aug 1985.

Saracevic, Tefko, et al. *Experiments on the Cognitive Aspects of Information Seeking and Retrieving.* Final Report for National Science Foundation Grant IST-8505411. Rutgers University, School of Communication, Information and Library Studies. Jan 1987.

U.S. Department of Health and Human Services, National Institutes of Health, National Library of Medicine. *Grateful MED™ User's Guide: Version 2.0.* 1986.

Arnold, Steve. "End Users: Old Myths and New Realities." *Proceedings of the Seventh National Online Meeting*, New York, 6–8 May 1986. Medford, N.J.: Learned Information, Inc. (1986): pp. 5 – 10.

Borgman, Christine L., Donald Case, and Charles T. Meadow. "Incorporating Users' Information Seeking Styles into the Design of an Information Retrieval Interface." *Proceedings of the 48th American Society for Information Science (ASIS) Meeting,* Las Vegas, Nevada, 20–24 Oct 1985. White Plains, N.Y.: Knowledge Industry Publications, Inc.(1985): pp. 324 – 330.

Curtis, Dade T. "Has the End User Been Forgotten? An Information Profile of Active Research Bioscientists." *The Value of Information: Collection of Papers presented at 6th Mid-Year Meeting*, Syracuse, N.Y., 19–21 May 1977. Syracuse University: American Society for Information Science (1977).

Obermeier, K. K. "Natural Language Front-Ends for Expert Systems." *Proceedings of the Fifth National Online Meeting,* New York, 10–12 Apr 1984. Medford, N.J.: Learned Information, Inc. (1984): pp. 265 – 272.

Proceedings of the Second Conference on Computer Interfaces and Intermediaries for Information Retrieval. Boston, Mass., 28–31 May 1986. Defense Technical Information Center and the Massachusetts Institute of Technology. AD-A174 000. (1986).

Santosuosso, J. "Requirements for Gateway Software for Libraries." *Proceedings of the Seventh National Online Meeting*, New York, 6–8 May 1986. Medford, N.J.: Learned Information, Inc. (1986): pp. 409 – 413.

Snow, Bonnie. "Self-Help Aids for End Users." *Proceedings of the Seventh National Online Meeting,* New York, 6–8 May 1986. Medford, N.J.: Learned Information, Inc. (1986): pp. 427 – 431.

Borgman, Christine L. "The User's Mental Model of an Information Retrieval System: Effects on Performance." Ph.D. diss., Stanford University, 1984.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

A193 362

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT "A" Approved for Public Release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) LMI-DL604R1	5. MONITORING ORGANIZATION REPORT NUMBER(S) DTIC-88/7 -TR-
6a. NAME OF PERFORMING ORGANIZATION Logistics Management Institute 6b. OFFICE SYMBOL (*If applicable*)	7a. NAME OF MONITORING ORGANIZATION Defense Technical Information Center
6c. ADDRESS (*City, State, and ZIP Code*) 6400 Goldsboro Road Bethesda, Maryland 20817-5886	7b. ADDRESS (*City, State, and ZIP Code*) Cameron Station Alex., VA 22304-6145
8a. NAME OF FUNDING / SPONSORING ORGANIZATION National Library of Medicine 8b. OFFICE SYMBOL (*if applicable*)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-85-C-0139
8c. ADDRESS (*City, State, and ZIP Code*) 8600 Rockville Pike Bethesda, MD 20894	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (*Include Security Classification*)
Capitalizing on Experience with Intelligent Gateway Software

12. PERSONAL AUTHOR(S)
Cynthia W. Shockley

13a. TYPE OF REPORT	13b. TIME COVERED	14 DATE OF REPORT (*Year, Month, Day*)	15. PAGE COUNT
Final	FROM _____ TO _____	January 1988	144

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (*Continue on reverse if necessary and identify by block number*)
FIELD	GROUP	SUB-GROUP	Intelligent gateway, DGIS, SearchMAESTRO, Micro-CSIN, Grateful MED

19. ABSTRACT (*Continue on reverse if necessary and identify by block number*)

The Defense Technical Information Center (DTIC) and National Library of Medicine (NLM) both developed gateway software to help users search, retrieve, and analyze information from different data systems, even when hardware and software incompatibilities exist. DTIC has two gateways: Defense Gateway Information System (DGIS) and Search Menu-Aided Easy Searching Through Relevant Options (SearchMAESTRO). NLM also has two: Micro-Chemical Substances Information Network (Micro-CSIN) and Grateful MED. Each has strengths and weaknesses; DTIC and NLM experience with them, along with commercial and academic developments, reveal opportunities for improvements.

DGIS software excels in processing results of a search, but offers little help in devising search strategies. SearchMAESTRO provides easier database searching, particularly for novices. DTIC software lacks guidance for choosing between gateways based on user need and experience.

With NLM's Micro-CSIN, users can search databases using scripts to define search strategies. The user interface is unwieldy but Micro-CSIN provides excellent assistance in composing a search strategy and selecting databases. Grateful MED enables users to review and incorporate terms and names from controlled vocabularies into a search strategy. Initial plans for NLM's Advanced, Biomedical Information Databand Exchange (ABIDE) will further the dissemination of biotechnology information. NLM has not yet adapted present gateways into ABIDE.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS		
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (*Include Area Code*)	22c. OFFICE SYMBOL

19. ABSTRACT *(Continued)*

We recommend that:

- DTIC enhance DGIS to enable all users to structure search strategies and select suitable databases, and adapt DGIS so that search terms used for one database can be used in several. Such improvements should enable DTIC to use DGIS to meet all user needs.
- NLM improve the "user-friendliness" of Micro-CSIN with online and window prompting techniques and modify Micro-CSIN to incorporate controlled vocabulary terms and names into a search strategy. NLM should decide how to incorporate Micro-CSIN and Grateful MED into ABIDE.
- DTIC and NLM adopt developments from the commercial and academic sectors such as an object-oriented gateway environment, develop expert systems to aid in creating more effective search strategies, and consider cooperative funding arrangements for future gateway development.

END

DATE

FILMED

DTIC

JULY 88